STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 145028

TO: Bao-Qun Li
Location: rem/3d24/3c18
Art Unit: 1648
Wednesday, February 23, 2005

Case Serial Number: 09/664363

From: **Deirdre Arnold**
Location: **Biotech-Chem Library**
REM 1A64
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

Search Notes

Please feel free to contact me if you have any questions or would like to amend the search.

Thank you for using STIC services.

Regards,
Deirdre Arnold

This Page Blank (uspto)

From:	Li, Bao-Qun
Sent:	Monday, February 14, 2005 8:42 AM
To:	STIC-Biotech/ChemLib

PLEASE DO THE SEQUENCE HOMOLOGY AND INTERFERENCE OF SEQ ID NO; 19 IN APPLICATION SN.
09,664,363. THANKS.
Bao Qun Li M.D
TC 1600
Art Unit 1648
Tel. 517-272-0904
REM, 3C18
Rm. 3D24

****
STAFF USE ONLY

Searcher:_____
Searcher Phone: 2-
Date Searcher Picked up:_____
Date Completed:_____
Searcher Prep/Rev. Time:___
Online Time:_____

****
Type of Search
NA Sequence: #_____
AA Sequence :#_____
Structure: #_____
Bibliographic:_____
Litigation:_____
Patent Family:_____
Other:_____

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other(Specify):_____

This Page Blank (uspto)

STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 145027

TO: Bao-Qun Li **Location: rem/3d24/3c18** **Art Unit: 1648** **Sunday, February 20, 2005** **Case Serial Number: 09/664363**	**From:** **Toby Port** **Location:** **Biotech-Chem Library** **REM-1A59** **Phone:** **571-272-2523** **toby.port@uspto.gov**

Search Notes

Examiner Li,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Toby Port

STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➤ *I am an examiner in Workgroup:* ☐   *Example: 1610*

➤ *Relevant prior art found, search results used as follows:*

☐ 102 rejection
☐ 103 rejection
☐ Cited as being of interest.
☐ Helped examiner better understand the invention.
☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

☐ Foreign Patent(s)
☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ *Relevant prior art not found:*

☐ Results verified the lack of relevant prior art (helped determine patentability).
☐ Results were not useful in determining patentability or understanding the invention.

Comments:

sira
Search and Information

From: Li, Bao-Qun
Sent: Monday, February 14, 2005 8:41 AM
To: STIC-Biotech/ChemLib

PLEASE DO THE SEQUENCE HOMOLOGY AND INTERFERENCE OF SEQ ID NO; 5 IN APPLICATION SN.
09,664,363. THANKS.
Bao Qun Li M.D
TC 1600
Art Unit 1648
Tel. 517-272-0904
REM, 3C18
Rm. 3D24

RECEIVED (STIC) FEB 14 2005

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: February 20, 2005, 22:06:50 ; Search time 4114 Seconds
(without alignments)
9822.958 Million cell updates/sec

Title: US-09-664-363-5
Perfect score: 834
Sequence: 1 AGAAAAACCAAACGTAACAC..........TCTCGGGATCTGTTTTCCCG 834

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	834	100.0	834	6	A28135	A28135 PT-NANBH mR
2	834	100.0	834	6	A32186	A32186 NANBHV PT p
3	834	100.0	834	6	AR144034	AR144034 Sequence
4	804.8	96.5	2790	6	A76569	A76569 Sequence 1
5	804.8	96.5	3372	6	A76594	A76594 Sequence 26
6	802.6	96.2	2116	6	A28155	A28155 PT-NANBH mR
7	802.6	96.2	2116	6	A32202	A32202 NANBHV PT p
8	802.6	96.2	2116	6	AR144050	AR144050 Sequence
9	801.6	96.1	2116	6	A76571	A76571 Sequence 3
10	748.2	89.7	831	6	HCVL32997	HCVL32997 Hepatitis
11	746.6	89.5	1595	14	HPCNS1SPF	M74809 Hepatitis C
12	746.6	89.5	9379	14	AF207768	AF207768 Hepatitis
13	745	89.3	9379	14	AF165051	AF165051 Hepatitis
14	745	89.3	9379	14	AF165052	AF165052 Hepatitis
15	745	89.3	9410	14	HPCK1R2	D50481 Hepatitis C
16	743.4	89.1	3296	14	AB008441	AB008441 Hepatitis
17	743.4	89.1	3296	14	AB008443	AB008443 Hepatitis
18	743.4	89.1	9386	14	AF165055	AF165055 Hepatitis
19	743.4	89.1	9386	14	AF165056	AF165056 Hepatitis
20	743.4	89.1	9418	14	HCVL32996	AJ132996 Hepatitis
21	743.4	89.1	1635	14	HPCNS1SPD	M74807 Hepatitis C
22	741.8	88.9	2672	14	HCVJK2	X61593 Hepatitis C
23	741.8	88.9	3296	14	AB008442	AB008442 Hepatitis
24	741.8	88.9	9373	14	AF208024	AF208024 Hepatitis
25	741.8	88.9	9573	14	AF207753	AF207753 Hepatitis
26	741.8	88.9	9585	14	AB049090	AB049090 Hepatitis
27	741.8	88.8	9585	14	AB049095	AB049095 Hepatitis
28	740.2	88.8	9379	14	AF207761	AF207761 Hepatitis
29	740.2	88.8	9379	14	AF207766	AF207766 Hepatitis
30	740.2	88.8	9379	14	AF207773	AF207773 Hepatitis
31	738.6	88.8	9410	14	HPCK1S2	D50485 Hepatitis C
32	738.6	88.6	9402	14	HCU45476	E05027 Hepatitis C
33	737	88.6	9431	14	HCU45476	U45476 Hepatitis C
34	737	88.4	2433	6	A48711	A48711 Sequence 49
35	737	88.4	2433	6	AR157351	AR157351 Sequence
36	737	88.4	2433	6	AX452798	AX452798 Sequence
37	737	88.4	2433	6	AX685050	AX685050 Sequence
38	737	88.4	2433	6	AX794892	AX794892 Sequence
39	737	88.4	2672	14	HCVJK5	X61595 Hepatitis C
40	737	88.4	9033	14	D89872	D89872 Hepatitis C
41	737	88.4	9369	14	AF165053	AF165053 Hepatitis
42	737	88.4	9426	14	AB049087	AB049087 Hepatitis
43	737	88.4	9547	14	AB049091	AB049091 Hepatitis
44	737	88.4	9587	14	AF333324	AF333324 Hepatitis
45	735.4	88.2	1195	14	HPCCOPR	L38318 Hepatitis C

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
A28135
LOCUS   A28135   834 bp   DNA   linear   PAT 07-JUN-1995
DEFINITION PT-NANBH mRNA fragment from patent GB2239245.
ACCESSION   A28135
VERSION   A28135.1 GI:1248620
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM synthetic construct
            other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 834)
AUTHORS
TITLE   Post-transfusional non-A, non-B hepatitis viral polypeptides
JOURNAL   Patent: GB 2239245-A 5 26-JUN-1991;
FEATURES   Location/Qualifiers
     source   1..834
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /clone="BR11"
     CDS   <1..>834
                     /codon_start=1
                     /transl_table=11
                     /product="PT-NANBH polyprotein"
                     /protein_id="CAA01927.1"
                     /db_xref="GI:1248621"
                     /translation="RKTKRNTNLRPQDVRFPGGGQIVGGVYLLPRRGPRLGVRATRKT
                     SERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPT
                     DPRRRSRNLGKVLDTLTCGFADSHGVHSARRSLRGARARALAHGVRVLEDGVNYATGN
                     LPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEYADMIMHTPGC
                     VPCVREGNSSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFSSAMYVGDLCGSV
                     FP"
ORIGIN

Qy   1 AGAAAAACCAAACGTAACACCTAACACCACCAACCTCCCCCCACAGGACGTCAGGTTCCCGGGGGCGGGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGAAAAACCAAACGTAACACCACCAACCTCCGCCCACAGGACGTCAGGTTCCCGGGCGTGGT 60

Query Match   100.0%; Score 834; DB 6; Length 834;
Best Local Similarity 100.0%;
Matches 834; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pred. No. 5.2e-193;

```
RESULT 2
A32186
LOCUS       A32186                   834 bp    RNA     linear   PAT 10-DEC-1996
DEFINITION  NANBHV PT polyprotein fragment.
ACCESSION   A32186
VERSION     A32186.1  GI:1926542
KEYWORDS    .
SOURCE      Non-A, non-B hepatitis virus
  ORGANISM  Non-A, non-B hepatitis virus
            Viruses; unclassified viruses.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   .
  JOURNAL   Patent: FR 2655990-A 5 21-JUN-1991;
            1. . 834
FEATURES             Location/Qualifiers
     source          1. .834
                     /organism="Non-A, non-B hepatitis virus"
                     /mol_type="unassigned RNA"
                     /db_xref="taxon:12440"
                     /clone="BR11"
                     /clone_lib="lambda gt11"
     CDS             <1. .>834
                     /codon_start=1
                     /product="partial NANBH-PT polyprotein"
                     /protein_id="CAA02090.1"
                     /db_xref="GI:1926543"
                     /translation="RKTKRNTNLRPDDVRFPGGGQIVGGVVLLPRGPRLGVRATRKT
                     SERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPT
                     DPRRSRNLGKVIDTLTCGFADSHGVHSARRSRLRGARALAHGVRVLEDGVNYATGN
                     LPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETADIMHTPGC
                     VPCVRENNSSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFSSAMYVGDLCGSV
                     FP"
ORIGIN
```

```
Query Match      100.0%;  Score 834;  DB 6;  Length 834;
Best Local Similarity   100.0%;
Matches  834;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
Pred. No. 5.2e-193;
```

```
Qy     61 CAGATCGTTGGTGGGAGTTTACCTGTTGCCGCGGCCGCAGGGCGGGGCCCCAGGGTTGGG    120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGATCGTTGGTGGGAGTTTACCTGTTGCCGCGGCCGCAGGGCGGGGCCCCAGGGTTGGG    120

Qy    121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTTGGTAGGGAAGGGACAAACCTTATCCC    180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTTGGTAGGGAAGGGACAAACCTTATCCC    180

Qy    181 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTCTATGTGGCAAC        240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTCTATGTGGCAAC        240

Qy    241 GAGGGCCATGGGGGTGGGGCAGGATGGCTCCCGTGTCACCCCGTGGGGGAAGGGC        300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGGGCCATGGGGGTGGGGCAGGATGGCTCCCGTGTCACCCCGTGGGGGAAGGGC        300

Qy    301 CCCAACTGACCCCCCGGGTAGGGTGCGTAATTTGGGTAAAAGTCATCGATACCCTCACATGC   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCCAACTGACCCCCCGGGTAGGGTGCGTAATTTGGGTAAAAGTCATCGATACCCTCACATGC   360

Qy    361 GGCTTCGGCCGACTCTCATGGGGGTTACATTCCGGTCGTCGTTGGGCGTCTCCCGTTGCC    420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCTTCGGCCGACTCTCATGGGGGTTACATTCCGGTCGTCGTTGGGCGTCTCCCGTTGCC    420

Qy    421 AGGGGCCCTGGGGCCATGGGGCGTTCTGGAGGACGGGCGGCTGAACTATGCAACAGGGAAT   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGGGCCCTGGGGCCATGGGGCGTTCTGGAGGACGGGCGGCTGAACTATGCAACAGGGAAT   480

Qy    481 TTACCCCGGTTGCTCTCTTTTCTCTATCTTCCTGGCTTTGCTGTCGTTGTTTGACCATTCCA   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTACCCCGGTTGCTCTCTTTTCTCTATCTTCCTGGCTTTGCTGTCGTTGTTTGACCATTCCA   540

Qy    541 GCTTCCCGCTTATGAAGTCGCAACGTGTCCCGGGATCCTCGGAACGATTGCTCC        600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCTTCCCGCTTATGAAGTCGCAACGTGTCCCGGGATCCTCGGAACGATTGCTCC        600

Qy    601 AACTCAAGCATCGTGTACGGAGACAGCGGACACATGATCATCGATACCCTCACATGC     660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AACTCAAGCATCGTGTACGGAGACAGCGGACACATGATCATCGATACCCTCACATGC     660

Qy    661 TGTGTCCGGGAGGGTAATTCCTCCCGCGCTGGGGTAGCGCTCACTCCCAGCGGCTCC     720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TGTGTCCGGGAGGGTAATTCCTCCCGCGCTGGGGTAGCGCTCACTCCCAGCGGCTCC     720

Qy    721 AAGGACGGCCACAGCATCCCACACTCGGACAATACGACGCCACGTCGATTGCTCGTTGGGGCG   780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AAGGACGGCCACAGCATCCCACACTCGGACAATACGACGCCACGTCGATTGCTCGTTGGGGCG   780

Qy    781 GCTGCCTTCTCGTCCGGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCCG      834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCTGCCTTCTCGTCCGGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCCG      834
```

```
RESULT 3
AR144034
LOCUS       AR144034       834 bp    DNA       linear   PAT 08-AUG-2001
DEFINITION  Sequence 5 from patent US 6210675.
ACCESSION   AR144034
VERSION     AR144034.1  GI:15105901
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   Highfield,P.Edmund., Rodgers,B.Colin., Tedder,R.Seton. and
            Barbara,J.Anthony.James.
  TITLE     PT-NANB hepatitis polypeptides
  JOURNAL   Patent: US 6210675-A 5 03-APR-2001;
FEATURES    Location/Qualifiers
  source    1..834
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

Query Match          100.0%;   Score 834;   DB 6;   Length 834;
Best Local Similarity 100.0%;   Pred. No. 5.2e-193;
Matches  834;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 AGRAAAAACCAAACGTAACACCAACCTCCCGCCCCACAGGACGTCAGTTCCCGGCGGTGGT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AGRAAAAACCAAACGTAACACCAACCTCCCGCCCCACAGGACGTCAGTTCCCGGCGGTGGT 60

Qy    61 CAGATCGTTGGTGGAGTTTACCTGTTGCGGCGCCCGCCCCCCAGGTTGGGTGTCCGCGCGG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CAGATCGTTGGTGGAGTTTACCTGTTGCGGCGCCCGCCCCCCAGGTTGGGTGTCCGCGCGG 120

Qy   121 ACTAGGAAGACTTCCGAGCGGGTCGGCAACCTCGTGGAAGGCGACAAACCTATCCCCAAGGCT 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTAGGAAGACTTCCGAGCGGGTCGGCAACCTCGTGGAAGGCGACAAACCTATCCCCAAGGCT 180

Qy   181 CGCCAGCCCCGAGGGCAGGCGCCTGGGCTCAGCCCGGGTACCCTTGGCCCTTCTATGGCAAC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CGCCAGCCCCGAGGGCAGGCGCCTGGGCTCAGCCCGGGTACCCTTGGCCCTTCTATGGCAAC 240

Qy   241 GAGGGCCATGGGGGTGGGGCAGGATGGCTCCTGTCACCCGTGGCTCCCGGCCTAGTTGGGGC 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGGGCCATGGGGGTGGGGCAGGATGGCTCCTGTCACCCGTGGCTCCCGGCCTAGTTGGGGC 300

Qy   301 CCCACTGACCCCCCGCGTAGGTCCCGTAATTTGGGTAAAGTCATCGATACCCCTCACATGC 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCCACTGACCCCCCGCGTAGGTCCCGTAATTTGGGTAAAGTCATCGATACCCCTCACATGC 360

Qy   361 GGCTTCGCCGACTCTCATGGGTACATTCCGCTGTCGCGCTCCCTTAGGGGCCTGCC 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGCTTCGCCGACTCTCATGGGTACATTCCGCTGTCGCGCTCCCTTAGGGGCCTGCC 420

Qy   421 AGGGCCCTGGCCATGGCCGTCGCGGCGTTCTTGGAGGACGCGGCGTGAACTATACCAACAGGGAAT 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AGGGCCCTGGCCATGGCCGTCGCGGCGTTCTTGGAGGACGCGGCGTGAACTATACCAACAGGGAAT 480

Qy   481 TTACCCGGTTGCTCTTTCTCTATCTTCTTGGCTTTGTCTTGTTGTTGACCATTCCA 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTACCCGGTTGCTCTTTCTCTATCTTCTTGGCTTTGTCTTGTTGTTGACCATTCCA 540

Qy   541 GCTTCCGCGTTATGAAGTGCGGCCAACGTGTCCGGGATCTACCCATGTCACGAACGATTGCTCC 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCTTCCGCGTTATGAAGTGCGGCCAACGTGTCCGGGATCTACCCATGTCACGAACGATTGCTCC 600

Qy   601 AACTCAAGACATCGTGTACGAGACACGGACACAGCACTCGGCGGACAATGCTGC 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AACTCAAGACATCGTGTACGAGACACGGACACAGCACTCGGCGGACAATGCTGC 660

Qy   661 TGTGTCCGGGGAGGGGTAATTCCTCCCGCGCTCCCGCTCACTCCCAGGCTCAACGAG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TGTGTCCGGGGAGGGGTAATTCCTCCCGCGCTCCCGCTCACTCCCAGGCTCAACGAG 720

Qy   721 AAGGACGACCCAGCATCCCCACTGCCAACTACGACGGCCAGCAGTCGATTGCTCGTTGGGCG 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AAGGACGACCCAGCATCCCCACTGCCAACTACGACGGCCAGCAGTCGATTGCTCGTTGGGCG 780

Qy   781 GCTGCCTTCTCGTTCCGCTATGTACGTGGGGGATCTCGTGTTTTCCCG 834
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCTGCCTTCTCGTTCCGCTATGTACGTGGGGGATCTCGTGTTTTCCCG 834
```

```
RESULT 4
A76569
LOCUS       A76569         2790 bp   DNA       linear   PAT 19-OCT-1999
DEFINITION  Sequence 1 from Patent WO9317110.
ACCESSION   A76569
VERSION     A76569.1  GI:6088480
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2790)
  AUTHORS   Rodgers,B.C. and Parker,D.
  TITLE     A RECOMBINANT HEPATITIS C VIRUS POLYPEPTIDE
  JOURNAL   Patent: WO 9317110-A 1 02-SEP-1993;
            WELLCOME FOUND (GB)
FEATURES    Location/Qualifiers
  source    1..2790
            /organism="unidentified"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32644"
  CDS       1..>2790
            /note="unnamed protein product; Fusion of non-structural
            and structural protein sequences of PT-NANH"
            /codon_start=1
            /protein_id="CAB58621.1"
            /db_xref="GI:6088481"
            /translation="MPDYSYRPTPIGPDPDPDSLSAEFTEVDGVRLHRYAPACKPLLREEV
            TFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSA
            SQLSGPSSKATYITQNDFPDADLIEANLLWRHEMGGDITRVESENKRVVILDSFDPLRA
            EEDEREVSVPAELILRKSKFPPAMPAWARPDYNPPLLESWKAPDYVPPVVHGCPLPPT
            KTPPIPPPRKRTVVLTESTVSSALAEAELAKTKARGSSSGPSAVDSGTATAPPDQSSDDGG
            AGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMSYTWTGALITPCAA
            EESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQILDDHYQDVLKEMKAK
            ASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAINHIRSVWEDLLEDTET
            PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSS
            YGFQYSPGQRVEFLVNAWKSKKTPMGFAYDTCFDSTVTENDIRVEESIYQCCDLAPE
            ARQAIRSITERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAVKKK
            KKTIRNTNLRPQDVRFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
            KARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKVID
            TLTCGFADLMGYIPLVGAPLGGAARARLAHGVRVLEDGVNYATGNLPGCSFSIFLLALL
            SCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEFADMIMHTPGCVPCVREGNSSRCWV
            ALTPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMYVGDLCGSVFPEFQLSAGRYGSF
            PGTRQEEPKTHSLQGNP"
ORIGIN

Query Match          96.5%;   Score 804.8;   DB 6;   Length 2790;
Best Local Similarity 99.5%;   Pred. No. 7e-186;
Matches  828;   Conservative   0;   Mismatches   2;   Indels   2;   Gaps   2;

Qy     4 AAAAACCAAAAGTAACACCAACCTCCCGCCCCACAGGACGTCAGGTTCCCGGCGGTGTCAG 63
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1876 AAAAACCAAAAGTAACACCAACCTCCCGCCCCACAGGACGTCAGGTTCCCGGCGGTGTCAG 1935

Qy    64 ATCGTTGGTGGAGTTTACCTGTTGCGGCGCGGAGGGGCCCCCAGGTTGGGTGTGCGCGGACT 123
         ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  1936 ATCGTTGGTGGAGTTTACCTGTTGCGGCGCCCCAGGTTGGGTGTGCGCGGACT 1995

Qy   124 AGGAAGACTTCCGAGCGGGTCGGCAACCTCGTGGAAGGCGACAAACCTATCCCCAAGGCTCGC 183
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1996 AGGAAGACTTCCGAGCGGGTCGGCAACCTCGTGGAAGGCGACAAACCTATCCCCAAGGCTCGC 2055

Qy   184 CAGCCCCGAGGGCAGGGCCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTTCTATGGCAACGAG 243
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```
Db      2056 CAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAACGAG 2115

Qy       244 GGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGCCCC 303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2116 GGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGCCCC 2175

Qy       304 ACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGCGGC 363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2176 ACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGCGGC 2235

Qy       364 TTCGCCGACTCTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTA-GGGGCGCTGCCAG 422
             ||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||
Db      2236 TTCGCCGAC-CTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTAGGGGGCGCTGCCAG 2294

Qy       423 GGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAATTT 482
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2295 GGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAATTT 2354

Qy       483 ACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCCAGC 542
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2355 ACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCCAGC 2414

Qy       543 TTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTCCAA 602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2415 TTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTCCAA 2474

Qy       603 CTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGTGTGTGCCCTG 662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2475 CTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGTGTGTGCCCTG 2534

Qy       663 TGTCCGGGAGGGTAATTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAA 722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2535 TGTCCGGGAGGGTAATTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAA 2594

Qy       723 GGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGC 782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2595 GGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGC 2654

Qy       783 TGCCTTCTCGTCCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCCG 834
             ||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db      2655 TGCCTTCTGCTCCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCCG 2706

RESULT 5
A76594
LOCUS       A76594                  3372 bp    DNA     linear   PAT 19-OCT-1999
DEFINITION  Sequence 26 from Patent WO9317110.
ACCESSION   A76594
VERSION     A76594.1  GI:6088505
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3372)
  AUTHORS   Rodgers,B.C. and Parker,D.
  TITLE     A RECOMBINANT HEPATITIS C VIRUS POLYPEPTIDE
  JOURNAL   Patent: WO 9317110-A 26 02-SEP-1993;
            WELLCOME FOUND (GB)
FEATURES             Location/Qualifiers
     source          1. .3372
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
     CDS             1. .>3372
                     /note="unnamed protein product; Encodes a fusion protein
                     which can be expressed in insect cells infected with the
                     baculovirus BHC-28"
                     /codon_start=1
                     /protein_id="CAB58632.1"
                     /db_xref="GI:6088506"
                     /translation="MPDYSYRPTIGPDPPSLSAEFTEVDGVRLHRYAPACKPLLREEV
                     TFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSA
                     SQLSGPSSKATYITQNDFPDADLIEANLLWRHEMGGDITRVESENKVVILDSFDPLRA
                     EEDEREVSVPAEILRKSKKFPPAMPAWARPDYNPPLLESWKAPDYVPPVVHGCPLPPT
                     KTPPIPPPRRKRTVVLTESTVSSALAELATKAFGSSGPSAVDSGTATAPPDQSSDDG
                     AGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMSYTWTGALITPCAA
                     EESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQILDDHYQDVLKEMKAK
                     ASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAINHIRSVWEDLLEDTET
                     PIDTTIMAKNEVFCVQPERGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSS
                     YGFQYSPGQRVEFLVNAWKSKKTPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPE
                     ARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAGIIIC
                     DECHSTDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEI
                     PFYGKAIPIETIKGGRHLIFCHSKKKCDELAAKLVGLGINAVAYYRGLDVSVIPASGD
                     VVVVATDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRG
                     RTGRGRRGIYRFVAAVKKKKKKTKRNTNLRPQDVRFPGGGQIVGGVYLLPRRGPRLGV
                     RATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRGSR
                     PSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGV
                     NYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETADMI
                     MHTPGCVPCVREGNSSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMYVG
                     DLCGSVFPEFQLSAGRYGSFPGTRQEPKTHSLQGNP"
ORIGIN

  Query Match             96.5%;  Score 804.8;  DB 6;  Length 3372;
  Best Local Similarity   99.5%;  Pred. No. 7e-186;
  Matches  828;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy         4 AAAACCAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCCGGGCGGTGGTCAG 63
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2458 AAAACCAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCCGGGCGGTGGTCAG 2517

Qy        64 ATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCGACT 123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2518 ATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCGACT 2577

Qy       124 AGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCTCGC 183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2578 AGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCTCGC 2637

Qy       184 CAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAACGAG 243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2638 CAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAACGAG 2697

Qy       244 GGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGCCCC 303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2698 GGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGCCCC 2757

Qy       304 ACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGCGGC 363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2758 ACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGCGGC 2817

Qy       364 TTCGCCGACTCTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTA-GGGGCGCTGCCAG 422
             ||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||
Db      2818 TTCGCCGAC-CTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTAGGGGGCGCTGCCAG 2876

Qy       423 GGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAATTT 482
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2877 GGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAATTT 2936

Qy       483 ACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCCAGC 542
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2937 ACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCCAGC 2996

Qy       543 TTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTCCAA 602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2997 TTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTCCAA 3056

Qy       603 CTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGTGTGTGCCCTG 662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3057 CTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGTGTGTGCCCTG 3116

Qy       663 TGTCCGGGAGGGTAATTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAA 722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3117 TGTCCGGGAGGGTAATTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAA 3176

Qy       723 GGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGC 782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 6

```
Db  3177 GGAGCGCCAGCATCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGGGTCGGC 3236

Qy   783 TGCCTTCTCGTCCCTATGTAGTTACGTGGGGGGGGGGGATCTCTGCGGGATCTGTTTTCCCG 834
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  3237 TGCCTTCTCGTCCCTATGTACGTGGGGGGGATCTCTGCGGGATCTGTTTTCCCG 3288
```

FEATURES   Location/Qualifiers
     source   1..2116
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="contig formed by cDNA clones"
     CDS   308..>2116
                     /codon_start=1
                     /transl_table=11
                     /product="PT-NANBH polyprotein N-terminus"
                     /protein_id="CAA01931.1"
                     /db_xref="GI:1248639"
                     /translation="MSTNPKPQRTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETAD
                     MIMHTPGCVPCVREGNSSRCWVALTPTLAAKDASIPPATIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHQTVQDCNCSIYPGHVSGHRNAWDMMMNWSPTAALVV
                     SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLVVMLLFAGVDGEPYTTGGTH
                     GRAAHGILTSLFTPGPAQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHRFNAS
                     GCSERMASCRPIDQFDQGWGPITYNESHGLDQRPYCWHYAPQPCGIVPALQVCGPVYC
                     FTPSPVVVGTTDRFGAPTYRWGENETDVLLNNTRPPRGNWFGCTWMNSTGFTKTCGG
                     PPCNIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWL"
```

ORIGIN

```
Qy     1 AGAAAAACCAAAGTAACACCAACCTCCCCCCACAGGAGTCAGTTCCCGGGCGGT  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   332 AGAAAAACCAAAGTAACACCAACCTCCCCCCACAGGAGTCAGTTCCCGGGCGGT  391

Qy    61 CAGATCGTTGGTGGAGTTTACTGTTGCCGGCGGCCCCAGGGGTTGGGGGTTGTCGCGCCGG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   392 CAGATCGTTGGTGGAGTTTACTGTTGCCGGCGGCCCCAGGGGTTGGGGGTTGTCGCGCCGG  451

Qy   121 ACTAGGAAGACTTCCGAGCGGTCCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   452 ACTAGGAAGACTTCCGAGCGGTCCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT  511

Qy   181 CGCCAGCCCGAGGGCCAGGGCGCCCTGGCGTCAGCCCCGGGTACCCTTGGCCCCTCTATGGCCAAC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   512 CGCCAGCCCGAGGGCCAGGGCGCCCTGGCGTACCCTTGGCCCCTCTATGGCCAAC  571

Qy   241 GAGGGCCATGGGGTGGGCCAGGATGGCCTCCGGGTCTCCGTGGCTCTCCGTGCCCGGGTCCCCGCGGGC  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   572 GAGGGCCATGGGGTGGGCCAGGATGGCCTCCGGGTCTCCGTGCCCGGGTCCCCGCGGGC  631

Qy   301 CCCACTGACCCCCGGGCGGTGTAGGTCGGTAATTTGGGTAAAGTCATCATCGATACCCTCACATGC  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   632 CCCACTGACCCCCGGGCGGTAGGTCGGTAATTTGGGTAAAGTCATCATCGATACCCTCACATGC  691
```

Query Match   96.2%; Score 802.6; DB 6; Length 2116;
Best Local Similarity 99.3%;
Matches 827; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

RESULT 7

FEATURES   Location/Qualifiers
     source   1..2116
                     /organism="Non-A, non-B hepatitis virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:12440"
     CDS   308..>2116
                     /codon_start=1
                     /product="partial NANBH-PT polyprotein"
                     /protein_id="CAA02094.1"
                     /db_xref="GI:1926563"
                     /translation="MSTNPKPQRTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETAD
                     MIMHTPGCVPCVREGNSSRCWVALTPTLAAKDASIPPATIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHQTVQDCNCSIYPGHVSGHRNAWDMMMNWSPTAALVV
                     SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLVVMLLFAGVDGEPYTTGGTH
                     GRAAHGILTSLFTPGPAQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHRFNAS
                     GCSERMASCRPIDQFDQGWGPITYNESHGLDQRPYCWHYAPQPCGIVPALQVCGPVYC
                     FTPSPVVVGTTDRFGAPTYRWGENETDVLLLNNTRPPRGNWFGCTWMNSTGFTKTCGG
                     PPCNIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWL"

ORIGIN

```
Qy   361 GGGCTTGCGCCGACTCGTCTCATGGGGGTACATTTCGCGTCGTGTGCTGCTTCGTGTGTCGTCGCGGGGTACTCGTTCCTGTTA-GGGGGGCTGC  419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||  ||||||||||
Db   692 GGCTTGCGCCGAC-CTCATCGGGGTACATTCGCGTCGTGTGCTGCTTCCTTTCCTCGCGGCGCGGGGGCTGC  750

Qy   420 CAGGGGCCCTGGCGCATGGGCGTCCGGGTTCTGGACGAGGACGAGGATTGCTC  479
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   751 CAGGGGCCCTGGCGCATGGGCGTCTCGGGAGGGGTGAACTATGCAACAGGAA  810

Qy   480 TTTTACCCCGGTTGCTCTTCTTCTCTATCTTCGCTTTGGCTTGCTTTTGTCTTCTGTTTGACATTCC  539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   811 TTTTACCCCGGTTGCTCTTCTCTATCTTCGCTTTGGCTTGCTTTTTGACATTCC  870

Qy   540 AGCTTCCGCTTATGAAGTGCGGCAACAGTGTCCGGGAGTTCTGGACGAGGACGAGGATTGCTC  599
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   871 AGCTTCCGCTTATGAAGTGCGGCAACAGTGTCCGGGAGTTCACGGAACGAGGATTGC  930

Qy   600 CAACTCAAGCACATCGTGTACGGACAGCAGCGACATGATCATGCACACCCCCGCGCCCC  659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   931 CAACTCAAGCACATCGTGTACGGACAGCAGCGACATGATCATGCACACCCCCGCGCCCC  990

Qy   660 CTGTGTCCCGGGAGGGGTAAATTCCTCCCGGCTCGTGGGTGGGTAGCCCTCCACTCCACCGACGCTCGCGGC  719
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   991 CTGTGTCCCGGGAGGGGTAAATTCCTCCCGGCTCGTGGGTGGGTAGCCCTCCACTCCACCGACGCTCGCGGC  1050

Qy   720 CAAGGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGTTGGGGTCGGGC  779
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1051 CAAGGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGTTGGGGTCGGGC  1110

Qy   780 GGCTGCCTTCTCGTCCGCTATGTAGGTACGTGGGGGGGGGGGATCTCTGCGGGATCTGTTTTCC  832
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1111 GGCTGCCTTCTCGTCCCTATGTAGGTACGTGGGGGGGGGGGATCTCTGCGGGATCTGTTTTCC  1163
```

Query Match   96.2%; Score 802.6; DB 6; Length 2116;
Best Local Similarity 99.3%;
Matches 827; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```
RESULT 8
LOCUS       AR144050     2116 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 21 from patent US 6210675.
ACCESSION   AR144050
VERSION     AR144050.1  GI:15105917
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2116)
  AUTHORS   Highfield,P.Edmund., Rodgers,B.Colin., Tedder,R.Seton. and
            Barbara,J.Anthony.James.
  TITLE     PT-NANB hepatitis polypeptides
  JOURNAL   Patent: US 6210675-A 21 03-APR-2001;
FEATURES             Location/Qualifiers
     source          1..2116
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match             96.2%;  Score 802.6;  DB 6;  Length 2116;
Best Local Similarity   99.3%;  Pred. No. 2.4e-185;
Matches  827;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;
```

[Sequence alignment block between Qy (query) and Db (database), bases 1–1163]

```
RESULT 9
LOCUS       A76571       831 bp    DNA     linear   PAT 19-OCT-1999
DEFINITION  Sequence 3 from Patent WO9317110.
ACCESSION   A76571
VERSION     A76571.1  GI:6088482
```

[Sequence alignment block between Qy (query) and Db (database), bases 660–1163]

```
KEYWORDS      unidentified
SOURCE        unidentified
  ORGANISM    unidentified
              unclassified.
REFERENCE     1 (bases 1 to 831)
  AUTHORS     Rodgers,B.C. and Parker,D.
  TITLE       A RECOMBINANT HEPATITIS C VIRUS POLYPEPTIDE
  JOURNAL     Patent: WO 9317110-A 3 02-SEP-1993;
              WELLCOME FOUND (GB)
FEATURES             Location/Qualifiers
  source      1..831
              /organism="unidentified"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32644"
  CDS         <1..>831
              /codon_start=1
              /product="PT-NANBH"
              /protein_id="CAB58622.1"
              /db_xref="GI:6088483"
              /translation="KTKRNTNLRPQDVRFPGGGQIVGGVYLLPRRGPRLGVRAATRKTS
              ERSQPGRQRPIPKARQPEGRAWAQPGYPWPLYGNGMGWAGWLLSPRGSRPSWGPTD
              PRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNL
              PGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETADMIMHTPGCV
              PCVREGNSSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMYVGDLSGSVF
              P"
              /note="Encodes PT-NANBH structural protein region"

ORIGIN
Query Match    96.1%;  Score 801.6;  DB 6;  Length 831;
Best Local Similarity    99.3%;  Pred. No. 4.3e-185;
Matches    826;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy      4 AAAACCAAAAGTAACACCAACCTTCCGCCCACAGAGACGTGAGGTTCCCGGGCGGTGGTCAG     63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAACCAAAACGTAACACCAACCTTCCGCCCCACACAGACGTGAGGTTCCCGGGCGGTGGTCAG     60

Qy     64 ATCGTTGGTGGAGTTTACCTGTTGCCGCGGCAGGGGCCCCCAGGTTGGGTGTGCCGCGACT     123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCGTTGGTGGAGTTTACCTGTTGCCGCGGCAGGGGCCCCCAGGTTGGGTGTGCCGCGACT     120

Qy    124 AGGAAGACTTCCGAGCGGTCGCAACCTCGTGGGAAGGCGGACGACAACCTATCCCCAAGGCTCGC     183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGAAGAGTTCCGAGCGGTCGCAACCTCGTGGGAAGGCGGACAACCTATCCCCAAGGCTCGC     180

Qy    184 CAGCCCGAGGGCAGGGGCCTGGGCCTCAGCCCGGGTACCCCTTGGCCCTCTATGGCAACGAG     243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGCCCGAGGGCAGGGGCCTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATGGCAACGAG     240

Qy    244 GGCATGGGGTGGGGCAGGAATGGCTGCTTCTGTCACCCGGGCCTGCTTCCGGGGGCCCC     303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCATGGGGTGGGGCAGGAATGGCTCCTGTCACCCGGGCCTCCTAGTTGGGGCCCC     300

Qy    304 ACTGACCCCCGGCGGTCGCGGTAATTGGGGTAAAGTCATCGATACCCTCACATGCGGC     363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACTGACCCCCGGCGGTAGGTCGCGGTAATTGGGGTAAAGTCATCGATACCCTCACATGCGGC     360

Qy    364 TTCGCCGACTCTCATGGGGTACAATTCCGCTCGTCGGCGGCTCGCCAG     422
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTCGCCGAC--CTCATGGGGTACAATTCCGCTCGTCGGCGGCTGCCCAG     419

Qy    423 GGCCCCTGCGCGCATGGGCGTCCGGGTTCTCGAGAGACAGGGCGGGAATTT     482
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 GGCCCCTGCGCGCATGGGCGTCCGGGTTCTCGAGAGCAGGGCGGGAATTT     479

Qy    483 ACCCGGTTGCTCTTTCCTTCTTCCTTGTGCTTGACCATTCCAGC     542
          ||||||||||||||||||||||||||||||||||||||||||||
Db    480 ACCCGGTTGCTCTTTCTCTATCTTCCTTGGCTTGACCATTCCAGC     539

Qy    543 TTCCGCTTATGAGTGCCGCAACGTGTCCCGGGATCTACCATCTACCATGCTCTCCAA     602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 TTCCGCTTATGAGTGCCGCAACGTTCACGAACGGATTGCTCTTTCCAA     599

Qy    603 CTCAAGCATCGTGTACGAGACGACGGACATGATCATGCACACCCCCGGGGTGTGTGCCCTG     662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 CTCAAGCATCGTGTACGAGAGACAGGCGGACAGAGGAAGCATGATCATGCACACCCCGGGGGTGTCGCTG     659

Qy    663 TGTCCGGAGGGTAATTCCTTCCCCCTCGTGCTGCTGGGTAGCGGTCACTCCACCCCAA     722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 TGTCCGGAGGGGTAATTCCTTCCCCTCGTGCTGCTGGGTAGCGGTCTCCACCCCAA     719

Qy    723 GGACGCCACGAGCATCCCACTGCGACAATACGAACGCCACCGTGCGGATTTGCTCGTTGGGGCGC     782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 GGACGCCACGAGCATCCCACTGCGACAATACGAACGCCACCGTGCGGATTTGCTCGTTGGGGCGC     779

Qy    783 TGCCTTCTCGTTCGGCTATGTTACGTGCGGGGGATCTCTGCGGGGATCTGTTTCCG     834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 TGCCTTCTCGTTCGGCTATGTGTACGTGCGGGGGATCTCTGCGGGGATCTGTTTCCG     831
```

```
RESULT 10
HCV132997           9533 bp    RNA    linear    VRL    04-MAY-1999
LOCUS         HCV132997              9533 bp    RNA    linear    VRL    04-MAY-1999
DEFINITION    Hepatitis C virus, complete genome, isolate HCV-AD78P1.
ACCESSION     AJ132997
VERSION       AJ132997.1  GI:4753720
KEYWORDS      complete genome; polyprotein.
SOURCE        Hepatitis C virus
  ORGANISM    Hepatitis C virus
              Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
              Hepacivirus.
REFERENCE     1
  AUTHORS     Rispeter,K., Lu,M., Lechner,S., Zibert,A. and Roggendorf,M.
  TITLE       Cloning and characterization of a complete open reading frame of
              the hepatitis C virus genome in only two cDNA fragments
  JOURNAL     J. Gen. Virol. 78 (Pt 11), 2751-2759 (1997)
  MEDLINE     9803183
  PUBMED      9367360
REFERENCE     2
  AUTHORS     Rispeter,K.
  JOURNAL     Thesis (1998) Universitaetsklinikum Essen, Institut fuer Virologie
REFERENCE     3 (bases 1 to 9533)
  AUTHORS     Rispeter,K.
  TITLE       Direct Submission
  JOURNAL     Submitted (07-MAR-1999) Rispeter K., Universitaetsklinikum Essen,
              Institut fuer Virologie, Hufelandstrasse 55, 45122 Essen, GERMANY
FEATURES             Location/Qualifiers
  source      1..9533
              /organism="Hepatitis C virus"
              /virion
              /mol_type="genomic RNA"
              /isolate="HCV-AD78P1"
              /db_xref="taxon:11103"
              /country="Germany"
              /note="Genotype 1b"
  CDS         341..9373
              /codon_start=1
              /product="polyprotein"
              /protein_id="CAB41951.1"
              /db_xref="GOA:Q9WIK8"
              /db_xref="UniProt/TrEMBL:Q9WIK8"
              /db_xref="GI:4753721"
              /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
              GVRATRKTSERSQPRGRRQPIPKARRPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRG
              SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
              GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGVYHVTNDCSNSSIVYETAD
              MIMHTPGCVPCVREDNFRCWVALTPTLAARNGSVPTTAIRRHVDLLVGAAAFCSAMY
              VGDLCGSVFLVSQLFTFSPRWHETVQDCNCSIYPGHVTGHRMAWDMMNWSPTTALVV
              SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIVTLLFAGVDGNTHIGGKQ
              AQATGGFVAWLAEGPSQEIQLINTNGSWHINRTALNCNDSLKTGFIALFYAHRFNSS
              GCPERMASCRPIDKFAQGWGPITYAEPGSSLDDRPYCWHYAPQPCGIVPASEVGGPVYC
              FTPSPVVVGTTDRSGVPTYRWGENETDVLLLNNTRPPQGNWFGCTWMNATGFTKTCGG
              PPCKIGGLGGNNTLTCPTDCFRKHPEATVTKCGSGPWLTPRCIVDYPYRLWHYPCTVNF
              TIFKVRMVVGGIEHRLSAACNWTRGERCDLEDRDRSELSPLLLSTTEWQIILPCSFTTL
              PALSTGLIHLHONTVDVQYLYGVGSVLVSFAIKWEYILLFFLLLADARVCACLWMMLL
              IAQAEAALENLVLLNAASLVGAHGILSFLVFFCAAWYVRGRLVPGAAYALYGVWPLLL
              LILALPPRAYAMDREMAASCGGAVFIGLALLTLSPYYVKVFLAKLIIWWLQYFITRAEAH
```

Left column

Db   LQVWIPPLNVRGGRDAIILLTCAVHPELIFPITKLLLAILGPLMVLQAGIIRAPYFVR
Qy   AQGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLVAVE
Qy   PVVFSDMETKIIIWGADTAACGDVILGLPVSARRGKEIFLGPADSLEGQGWRLLAPIT
Db   AYSQQTRGLLGCIIISLTGRDKNQVEGEVQVVSTATQSFLATCVNGWVTVFHGAGSK
Qy   TLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVSRHADVIPVRRRG
Db   DSRGSLLSPRPVSYLKGSSGGPLLCPSGHVAGIFRAAVCTRGVAKAVDFVPVESMETT
Qy   TRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
Db   SFGAYMSKAHGVDPSIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
Qy   TDSTLGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
Qy   AIPIEAIRGGRHLIFCHSKKKCDELAAQLSSLGVNAVAYYRGLDVSVIPTSGDVVVA
Db   TDALMTGFTGDFDSVIDCNTCVTQTVFDSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
Qy   KRGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQ
Db   DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
Qy   LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPVTKYIMACMSADLEVVTSTWVLVGGVLA
Db   ALAAYCLTTGSVVIVGRILSGKPAIIPDREVLYRFDFDMEECASHLPYIEQGMQLAE
Qy   QFKQKALGLLQTATKQAEAAAPVVESKWQALEAFWANDMWNFISGVQYLAGLSTLPGN
Qy   PAIASLMAFTAISSPLTTHQHTLLFNILGGWVAAQLAPPSAASAPVGAGIAGAAVGSI
Db   GLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAA
Db   KLRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
Qy   RLHQWISEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVFRFSCQRGYK
Qy   GVWRGDGIMQTTCPCGAQITHGVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSP
Db   APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHR
Db   YAPACPLLIRDEVTFQVGLNQYLVGSQLPCEPEPDVTVLTSMLTDPSHITAEAAKRRL
Qy   ARGSPPSLASSSASQLSAPSLKATCTTHRDSPDADLIEANLLWRQEMGGNITRVESEN
Db   KVVILDSFDPLRAEBEDEREISVPAEILRKSRSFPRAMPIWARPDYNPPLLESWKDPDY
Db   VPPVVHGCPLPPTRAPPIPPPRKKRTVVLTESTVSSALAELATKTFGSSSSSAVDSGT
Qy   ATAPPDQPSDGSDTGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEASEDVVCCSMS
Qy   YTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRKQKKVTFDRLQVLD
Db   DHYRDVLKEMKAKASTVKARLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAINHI
Db   NSVWEDLLEDTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
Qy   VVSTLPHAVMGSSYGFQYSPGQRVEYLVKAWKSSKNPMGFAYDTRCFDSTVTENDIRV
Qy   EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
Db   TCYLKASAACRAAKLQDCTMLVCGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPG
Db   DPPQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
Qy   GNIIMYAPTLWARMVLMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
Qy   GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSRGGRAATCGKYL
Db   FNWAVRTKLKLTPIPAASRLDLSGWFVAGYSGGDIYHSLSPARPRMWFMWLLLLSVGV
Db   GIYLLPNR"

ORIGIN

Query Match   89.7%; Score 748.2; DB 14; Length 9533;
Best Local Similarity   95.2%;
Matches 793; Conservative   0; Mismatches   38; Indels   2; Gaps   2;
                                   Pred. No. 4.6e-172;

Qy   1 AGAAAAACCAAAGTAACCAACCTTCCGCGCGGTCTCGTCTCTTGGTTTGGGTTCTTCGTT   60
Db   365 AGAAAAACCAAAGTAACACCAACCGCGCGCCCACCAGACGTCAAGTTCCCGGGTGGTCGGT   424

Qy   61 CAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCGGGCCAGGGGCGCCCCCAGGTTGGGGCG   120
Db   425 CAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCGGGCCAGGGGCGCCCCCAGGTTGTGCGCCG   484

Qy   121 ACTAGGAGGACTTCCGAGCGGTCGCAACCTCGTCGGCGGGAGGGCGACAAACCTTATCCCCAAGGCT   180
Db   485 ACTAGGAGGACTTCCGAGCGGTCGCAACCTCGTCGGCGGGAGGGCGACAAACCTTATCCCCAAGGCT   544

Qy   181 CGCCAGCGCCGCGAGGGCCAGGGCGCCTTGGGCCTCAGCGCCGGGTACCCTTGGGCCGCCAAC   240
Db   545 CGCCGCGGCCGCGAGGGCGGGCCTTGGGCCTCAGCGCCGGGTACCCTTGGGCCGCCCTCTATGCCGCAAC   604

Qy   241 GAGGGCCATGGGGGTTGGCAGGAGGATGGCTCTCCTGTCACCGGGCGTCCCGGCCGCCGCCGT   300
Db   605 GAGGGCCCTGGGGGTTGGCAGGAGGATGGCTCTCCTGTCACCGGGCGTCCCGGCCCTTAGTTGGGGC   664

Qy   301 CCCACTGACCCCCGGCGGTTCGGGTAAAGTCATCGATTACCCTTCACATGC   360
Db   665 CCCACAGACCCCCGGCGTTAGGGTCGGGTAAGGTCATCGATACCCTTCACATGC   724

Qy   361 GGCTTCGGCGACTTCTTCATCGGGGGTTACATGGGGGTTGGGGTTGCCGGGGCTTGC   419
Db   725 GGCTTCGGCGAC-CTCATCGGGGGTTACATTCCGGCGTCGGCGCCCCTTAGGGGGCACTGC   783

Qy   420 CAGGGGCCCTGGGCGGCCATGGCCGTGCGGTCCGGGTTCTTGGAGGACCGGCGGCGTGAAACTATGCAACAGGGAA   479

Right column

RESULT 11

HPCNS1SPF
LOCUS   HPCNS1SPF   1595 bp ss-RNA   linear   VRL 02-AUG-1993
DEFINITION Hepatitis C virus core protein (NS1) gene.
ACCESSION   M74809
VERSION   M74809.1 GI:329787
KEYWORDS   NS1 protein; core protein; envelope-associated protein.
SOURCE   Hepatitis C virus
  ORGANISM Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1 (bases 1 to 1595)
  AUTHORS   Cha,T.A., Beall,E., Irvine,B., Kolberg,J., Chien,D., Kuo,G. and
            Urdea,M.S.
  TITLE   At least five related, but distinct, hepatitis C viral genotypes
            exist
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (15), 7144-7148 (1992)
  MEDLINE   92357788
   PUBMED   1323128
COMMENT   Original source text: Hepatitis C virus (individual_isolate I15)
            cDNA to genomic RNA.
FEATURES   Location/Qualifiers
     source   1..1595
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="I15"
                     /db_xref="taxon:11103"
ORIGIN

Query Match   89.5%; Score 746.6; DB 14; Length 1595;
Best Local Similarity   95.1%;
Matches 792; Conservative   0; Mismatches   39; Indels   2; Gaps   2;
                                   Pred. No. 1.2e-171;

Qy   1 AGAAAAACCAAAGTAACCAACCTTCCGCGCGGTCTCGTCTCTTGGTTTGGGTTCTTCGTT   60
Db   266 AGAAAAACCAAAGTAACACCAACCGCGCGCCCACCAGACGTCAAGTTCCCGGGTGGTCGGT   325

Qy   61 CAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCGGGCCAGGGGCGCCCCCAGGTTGGGGCG   120
Db   326 CAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCGGGCCAGGGGCGCCCCCAGGTTGTGCGCCG   385

Qy   121 ACTAGGAGGAGACTTCCGCGCGGTCGCCAACCTCGTCGGCGGGAGGGCGACAAACCTTATCCCCAAGGCT   180
Db   386 ACTAGGAGGAGACTTCCGAGCGGTCGCAACCTCGTCGGCGGGAGGGCGACAAACCTTATCCCCAAGGCT   445

RESULT 12

LOCUS   AF207768   9379 bp   RNA   linear   VRL 27-APR-2000
DEFINITION Hepatitis C virus strain MD27 complete genome.
ACCESSION   AF207768
VERSION   AF207768.1 GI:7650253
KEYWORDS   .
SOURCE   Hepatitis C virus
  ORGANISM Hepatitis C virus
            viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1 (bases 1 to 9379)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Miyasaka,Y., Marumo,F. and
            Sato,C.
  TITLE   Characteristics of hepatitis C viral genome associated with disease
            progression
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 9379)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Miyasaka,Y., Izumi,N. and
            Sato,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku,
            Tokyo 113-8519, Japan
FEATURES   Location/Qualifiers
     source   1..9379
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="MD27"
                     /db_xref="taxon:11103"
                     /note="isolated from a patient with liver cirrhosis and
                     hepatocellular carcinoma (HCC-7)"
     CDS   330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF65958.1"
                     /db_xref="GI:7650254"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARRPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGVYHVTNDCSNSSIVYEAAD
                     MIMHTPGCVPCVREGNSSRCWVALTPTLAARNSSIPTTIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRYETVQDCNCSIYLGHVTGHRMAWDMMMNWSPTTALVV
                     SQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGRTRVMGGQA
                     AQHTRGASFSSPGAAQKIQLITNGSWHINRTALNCNDSLNTGFLAALFYAHKFNGS
                     GCPERMASCRRIDKFAQGWGPITHVPNISDQRPYCWHYAPQPCGIVPAAEVCGPVYC
                     FTPSPVVVGTTDRSGAPTYNWGENETDVLLNNTRPPLGNWFGCTWMNSTGFTKTCGA
                     PPCNIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TVFKVRMYVGGVEHRLKAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PALSTGLIHLHQNIVDVQYLYGVGSAVVSVIRWEYVLLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVILNAASVAGAHGILSFLVFFCAAWYIKGKLVPGAAYALYGAWPLLL
                     LLLALPPRAYAMDREMAASCGGVFIGLALLTLSPHYKVFLAQLIWWLQYFITRAEAH
                     LQVVPLNVRGGRDAIILLTCAAHSELVFDITKILLAILGPLMVLQAGITRVPYFVR
                     AQGLIRVCMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRHWAHAGLRDLAVAVE
                     PVVFSDMETKIITWGADTAACGDILLGLPVSARRGREILLGPADRFGEQGWRLLAPIT
                     AYSQQTRGLIGLIGIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVFHGAGTK
                     TLAGPKGPITQMYTNVDQDLVGWQAPGARSWTPCTCGSSDLYLVTRHADVIPVRRRG
                     DSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGVDPNIRTGVRTITTGSVTHYTSTYGKFLADGGCSGGAYDIIMCDECHS
                     TDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
                     AIPIETIKGGRHLIFCHSKKKCDELAAKLSALGINAVAYYRGLDVSVIPTSGDVVVVA
                     TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTPQDAVSRSQRRGRTGRG
                     RRGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTPLLYRLGAVQNEIITHPITKYIMTCMAADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGKPAVIPDREVLYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWRALEAFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTAISSPLTTQHTLLFNILGGWVAAQLAPPSASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGEIPSAEDMVNLLPAILSPGALVVGVVCAA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPENDAAARVTQILSSLTIQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYK
                     GVWRGDGVMQTTCPCGAQITGHVKNGSMRIVGPKACSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTELDGVRLHR
                     YAPACRPLLREEVTFQVGLNQYLVGSQLPCEPEPDVTVLTSMLTDPSHITAEAARRL
                     ARGSPPSLASSASQLSAPSLKATCTTHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVVILDSFDPLRAEEDEREVSVAAEILRKSRRFPPALPIWARPDYNPPLLESWKGPDY
                     PPPVVHGCPLPPTKAPPIPPPRRKRTVVLTESTVSSALAELATKTFGSSGSSAADGT
                     ATAPPDQPLDDGTGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMS
                     YTWTGALITTPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKAKLLSIEEACRLTPPHSAXSKFGYGAKDVRSLSSKAVNHI
                     HSVWKDLLEDTETPIDTTIMAKNEVFCVQPEKGGRKAARLIVFPDLGVRVCEKMALYD
                     VVSTLPQAVMGSSYGFQYSPGHRVEFLVNAWKSKKNPMGFAYDTRCFDSTVTESDIRV
                     EESIYQCCDLAPEARQVIRSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
                     TCYLKASAACRAAKLQDCTMLVCGDDLVVICESAGTQEDEASLRVFTEAMTRYSAPPG
                     DPPRPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTEPVNSWL
                     GNIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGATYSIEPLDLPQIIQRLH
                     GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYL
                     FNWAVRTKLKLTPIPAASQLDLSGWFVAGYSGGDIYHSLSHARPRWFLWCLLLLSVGV
                     GIYLLPNR"

ORIGIN
        Query Match   89.5%; Score 746.6; DB 14; Length 9379;
        Best Local Similarity 95.1%; Pred. No. 1.1e-171;
        Matches 792; Conservative   0; Mismatches   39; Indels   2; Gaps   2;

Qy   1 AGAAAAAACCAAACGTAACACCAACCTTCCGCCACCAGGACCTCAGGTTCCCGGGGGTGGT 60
            |||||||||| ||||||||||||||||||||| ||||||| |||||||||||||| ||||||
Db   354 AGAAAAAACCAAACGTAACACCAACCGCCCACCAGGAAGACTTCAAGTTCCCGGGTGGT 413

Qy   61 CAGATCGTTGGTGGAGTTTACCTTGTTGCCGCGCGCCCAGGGCGGCCCCAGTTGGGGTTGGGTGTGCGCGGGGT 120

RESULT 13

```
LOCUS       AF165051    9379 bp    RNA    linear    VRL 04-NOV-1999
DEFINITION  Hepatitis C virus strain MD4-1 complete genome.
ACCESSION   AF165051
VERSION     AF165051.1  GI:5918940
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9379)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Maekawa,S.Y., Miyasaka,Y.,
            Tazawa,J., Izumi,N., Marumo,F. and Sato,C.
  TITLE     Time-related changes in full-length hepatitis C virus sequences and
            hepatitis C virus activity
  JOURNAL   Virology 263 (1), 244-253 (1999)
   MEDLINE  20013335
   PUBMED   10544098
REFERENCE   2  (bases 1 to 9379)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Maekawa,S., Miyasaka,Y.,
            Sakamoto,N., Fukuma,T., Tazawa,J., Izumi,N., Marumo,F. and Sato,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku,
            Tokyo 113-8519, Japan
FEATURES             Location/Qualifiers
     source          1..9379
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="MD4-1"
                     /db_xref="taxon:11103"
                     /note="from a patient with persistently normal serum ALT
                     values; patient 4, point 1"
                     /genotype: 1b"
     CDS             330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAD56186.1"
                     /db_xref="GI:5918941"
                     /translation="MSTNPKPQRKTRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEAAD
                     VIMHTPGCVPCVRENNSSRCWVALTPTTLARNASIPTTTIRRHVDLLVGTAAFCSAMY
                     VGDLCGSVFLISQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVV
                     SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMAGNWAKVLIVMLLFAGVDGNTYTTSGGV
                     GRQTGLTSLFSVGASQKIQLINTNGSWHINRTALNCNDSLQTGFLAALFYRHRFNAS
                     GCPGRNASCRPIIDKFAQGWGPITYAEPSSDQRPYCWHYAPRPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRSGVPTYWGENKTDVLLNNTRPPQGNWFGCTWMNATGFTKTCGG
                     PPCHIGGVGNNTLICPTDCFRKHPEATYSRCGSGPWLTPRCLVHYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKWEYVLLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVVLNAASVARAHGILSFLVFFCAAWYIKGKLVPGAAYAIYGVWPLLL
                     LLLALPFSRAYAMDREMAASCGGAVFVGLALLTLSPYYKVFLARLMWWLQYFITRAEAH
                     LQVWIPPLNVRGGRDAIILLACAVHPELIFDITKLLLSILGPLMVLQASIVRVPYFVR
                     AQGLIRACMLVRKAAGGHYAQMVLLRLAAWLGTYVYDHLTPLRDVWAHAGRDLAVAVE
                     PVVFSDMETKVITWGADTAACGDIISGLPVSARRGREILLGPADSYEGQWRILAPIT
                     AYSQQTRGLLGCIISTLSTLTGRDKNQVEGEVQVVSTATQSFLATCVNGACWTVFHGAGSK
                     TLAGPKGPIYTQMYTNVDQDLVGWQAPGARSLITPCTCGSSDLYLVTRHADVIPVRRRG
                     DTRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVEAMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGTEPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
                     AIPIEVIKGGRHLIFCHSKKKCDELAAKLSALGLNAVAYYRGLDVSVIPTSGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RRGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYLNTPGLPVCQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGRPAVIPDREVLYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWRALEAFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSI
                     GLGKVIVDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALVVGVVCAA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
                     RLHQWINEDCSTPCSGSWLRDWDWICTVLSDFKTWLKAKLMPLLPGLPFLSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEVEVTRVGDFHYVTGMTTDNIKCPCQVPAPEFFTEVDGVRLHR
                     YAPACKPLLREEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAARRL
                     ARGSPPSLASSSASQLSAPSLKATCTTNHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVVILDSFDPLRAEEDEREPSVPAEILRKSRKFPMAMPVWARPDYNPPLLESWKDPDY
                     VPPVVHGCPLPPTKAPPIPPPRRKRTVVLTESTVSSALAELATKFGSSGSSAVDSGT
                     ATAPPDQASSDDGKDGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEASEDVVCCSMS
                     YTWTGALITPCAAEESKLPINSLSNSLLRHHNMVYATTSRSASLRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKAKLLSIEEACKLTPPHSAKSKFGYGAKDVRNLIGKAVNHI
                     RSVWKDLLEDTETPIDTTVMAKNEVFCIQPEKGGRKPARLIVFPDLGVREVKEKMALYD
                     VVSTLPQAVMGSSYGFQYSPKQRVEFLVNAWKSKKCPMGFSYDTRCFDSTVTENDIRV
                     EESIYQCCDLAPEARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSAPPG
                     TCYLKASAACRAAKLQDCTMLVNGDDLVICESAGTQEDAANLRVFTEAMTRYSAPPG
                     DPPQPAYDLELITSCSSNVSVAHDASGKRVYLTRDPTTPFARAAWETARHTPVNSWL
                     GNIIMYAPTLWARVLMTHFFSILLAQEQLEKALDCQIYGAHYSIEPLDLPQIERLH
                     GLSAFSLHSYSPGEINRVAACLRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYL
                     FNWAVRTKLKLTPIPAASQLDLSSWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGV
                     GIYLLPNR"
ORIGIN
```

Alignment:

```
Db   414 CAGATCGTTGGTGGAGTTTACCTGTTGCCGGCGCGGCCCAGGGCCCCCCAGGTTGGGTGTGCGCGCG  473
Qy   121 ACTAGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT  180
         |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   474 ACTAGGAAGAGACTTCTGAGCGGTTGCGACAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT  533
Qy   181 CGCCCAGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGGCAAC  240
         |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 CGCCCAGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAC  593
Qy   241 GAGGGCCATGGGGTGGGGCAGGATGGCTCCTGTCGTCGTGCTGTGCTCCCCCTTCTATGGGC  300
         ||||| |||||||||||||||||||| |||||||||
Db   594 GAGGGCTCTGGGGTGGGGCAGGATGGCTCCTGTCACCCCGGGCTCCCGGCCCTCTATGGGC  653
Qy   301 CCCACTGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAAGTCATCCTATCCCTCACATGC  360
         |||||| |||||||||||||||||||  |||||||||||||||||||||| ||||||||
Db   654 CCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGATACCCCTCACATGC  713
Qy   361 GGCTTCGCCGACTTCATGGGGTACATTCCGCTCGTCGGCGCTCCCCTTAGG--GGCGCTGC  419
         |||||||||||  ||||||||||||||||||||||||||||||||||||||  |||||||
Db   714 GGCTTCGCCGAC--CTCATGGGGTACATCCCGCTCGTCGGCGCCCCCCTAGGAGGCGGTGC  772
Qy   420 CAGGGGCCCTGCGGGCCATGGGGGTCTCGGGAGGAGGGGCTGAACTATGCAACAGAGGGAA  479
         ||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||
Db   773 CAGGGGCCCTGCAGGCGTCCGGGGTCTCGGGAGGAGGGGCTGAACTATGCAACAGAGGGAA  832
Qy   480 TTTACCCCGGTCGCTTTCTTCCTATCTCCTTGGCTTGGCCTTGCTCGTTGACCATTCC  539
         ||| |||||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db   833 TCTGCCCCGGTTGCTCTTTTCTCTATCTCCTTGGCTTGGCCTTGCGTCTTGGACCATTCC  892
Qy   540 AGCTTCCGCTTATGAAGTGCGCGAACGTGTCCGGGACTCACGAAGGATTGCTC  599
         ||||||||||||||||| |||| |||||||||||||||||||| |||||||||
Db   893 AGCTTCCGCCTATGAAGTGCGCAACGTGTCCGGGGTGTTACCACGTCACGAAGACTCTC  952
Qy   600 CAACTCAAGCATCGTGTACGACAGACAGGGCCATGAGTCAACATGATCGTGTGTGCC  659
         |||||||||||||||| |||||||||| |||||||||||||||||| ||||||||||
Db   953 CAACTCAAGCATTGTGTTATGAGGGACATGAGTCAACATGATGACCCCGGGTGTCC  1012
Qy   660 CTGTGTCCGGGAGGGTAATTCCTCCCCGGGACTCACTCCACTCCACGCCTGCGGGC  719
         ||||||||||| |||||||||||||||||| |||||||||||||||||||||||||
Db  1013 CTGCGTCCGGGAGGGCCAATTCCTCCCCCGGGACTCACGCCGGCCTCGTTGCCCGGGC  1072
Qy   720 CAAGGGACGGCCGCCAGCATCCCCACTCGACGGCACGTCGGATTTGCTCGTTCGGGC  779
         ||| |||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db  1073 CACGAATCGCAGCATCCCCGTTACGACAATACCGACGCGCCATGTCGTTGGGGC  1132
Qy   780 GGCTGCCTTCTCCTCGCGGCCTATGTAGTAGCGTGGGGGGGGATCTCTGCGGGATCTGTTTTCC  832
         |||| |||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  1133 AGTGCGTTTCTGCTCGCCGTATGTACGTGCGGGGGGGGATCTCTGCGCTATCTGTTTTCC  1185
```

RESULT 14

```
LOCUS       AF165052                9379 bp    RNA     linear   VRL 04-NOV-1999
DEFINITION  Hepatitis C virus strain MD4-2 complete genome.
ACCESSION   AF165052
VERSION     AF165052.1  GI:5918942
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9379)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Maekawa,S.Y., Miyasaka,Y.,
            Tazawa,J., Izumi,N., Marumo,F. and Sato,C.
  TITLE     Time-related changes in full-length hepatitis C virus sequences and
            hepatitis activity
  JOURNAL   Virology 263 (1), 244-253 (1999)
  MEDLINE   20013325
   PUBMED   10544098
REFERENCE   2  (bases 1 to 9379)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Maekawa,S., Miyasaka,Y.,
            Sakamoto,N., Fukuma,T., Tazawa,J., Izumi,N., Marumo,F. and Sato,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku,
            Tokyo 113-8519, Japan
FEATURES             Location/Qualifiers
     source          1..9379
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="MD4-2"
                     /db_xref="taxon:11103"
                     /note="from a patient with persistently normal serum ALT
                     values; patient 4, point 2"
                     /genotype:1b"
     CDS             330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAD56187.1"
                     /db_xref="GI:5918943"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEAAD
                     VIMHTPGCVPCVRESNSSRCWVALTPTLAARNASVPTTTIRRHVDLLVGTAAFCSAMY
                     VGDLCGSVFLISQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVV
                     SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMAGNWAKVLIVMLLFAGVDGETHTTGGAV
                     SRTTYRFASLFSXGASQKIQLIVTNGSWHINRTALNCNDSLQTGFLAALFYTHRFNAS
                     GCPGRMASCRPIDKFAQGWGPITYAEHDSPDDRPYCWHYAPKPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRSGVPTYTWGENKTDVLLLNNTRPPQGNWFGCTWMNATGFTKTCGG
                     PPCHIGGVGNNTLICPTDCFRKHPEATYSKCGSGPWLTPRCIVDYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKWEYVLLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVLNAASVAARGILPGILSPLVFPCAAWYIXGKLVPGAAYAIYGMVEA
                     LLLALPSRAYAMDREMAASCGGVFVGLALLSPYYKVFLARLMWWLQYFITRAEAH
                     LQVWVPPLNVRGGRDAIILLACAVHPELIFDITKLLLSILGPLMVLQASIVRVPYFVR
                     AQGLIRACMLVRKAAGHYAQMVLVRLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVE
                     PVVFSDMETKVITWGADTAACGDIISGLPVSARRGREIILGPADNYEGCGWVFHGAGSK
                     AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGACWTVFHGAGSK
                     TLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
                     DTRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVERAMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
                     AIPIEVIKGGRHLIFCHSKKKCDELAAKLSALGLNAVAYYRGLDVSVIPTSGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RRGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAVLNTPGLPVCQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKFIMACMSADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGRPAVIPDREVLYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWRALEAFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALVVGVVCA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTTQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLSDFKTWLQSKLLPRLPGLPFLSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRKTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTNDNIKCPCQVPAPEFFTEVDGVRLHR
                     YAPACKPLLREEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAARRL
                     ARGSPPSLASSASQLSAPSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVILLDSFDPLRAEEDEREPSVPAEILRKSRKFPMAMPVWARDYNPPLLESWKDPDY
                     VPPVVHGCPLPPPRKTPVPPPRRKRTVVLTESTVSSALAELATKTFGSSGSSAVDSGT
                     ATAPPDQASSDDGDKGSDVESYSSMPPLEGEPGDPLSDGSWSTVSEEASEDVVCCSMS
                     YTWTGALITPCAAEESKLPINSLSNSLLRHHNMVYATTSRSASLRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKAKLLSIEEACKLTPPHSAKSKFGYGAKDVRNLTGKAVNHI"
```

Query Match 89.3%; Score 745; DB 14; Length 9379;
Best Local Similarity 95.0%; Pred. No. 2.8e-171;
Matches 791; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

```
Qy     1 AGAAAAACCAAAACGTAACCAACCAACCTCCGCCCCACAGGACGTCAGGTTCCGGGCGGT    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   354 AGAAAAACCAAAACGTAACCAACCAACCTCCGCCCCACAGGACGTCAAGTTCCGGGCGGT   413

Qy    61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG   473

Qy   121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   474 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT   533

Qy   181 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAC   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 CGCCGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAT   593

Qy   241 GAGGGCCATGGGGTGGGCAGGAATGGCTCCTGTCACCCCTCCCGGGCTTAGTTGGGGGC   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 GAGGGCCATGGGGTGGGCAGGTTGGCTCCTGTCACCCCTCCCGGGCTTAGTTGGGGGC   653

Qy   301 CCCACTACTGACCCCCGGGCGTAGGTCCCGGTAATTTGGGTAAAGTCATCGGATACCCTC   360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   654 CCCACTGACCCCCGGGCGTAGGTCGCGTAATTTGGGTAAGGTCATCGATACCCTCACATGC   713

Qy   361 GGCTTCCGCGACTCTCATGGGGTACATTCGGCTCCTCGGCGTCCGGGCTCCCTTA-GGGGC   419
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||| |||||||
Db   714 GGCTTCGGGCGAC--CTCATGGGGTACATTCCGGCTCCTCGTCGCCCCGGGCTCCCTGC   772

Qy   420 CAAGGGCCCTGGCGCGTCCCGGGTTCTCGGAGGACGCGCGGGCTGAAACTATGGCACAGG   479
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   773 CAGGGGCCCTGGCGCGTCCCGGGTTCTCGGAGGACGCGCGGGCTGAAACTATGGCACAGG   832

Qy   480 TTTACCCGGTTGCTCTTCTTTCTATCTTCCTTCGGCTTTGGCTTTGTTGACCATTCC   539
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   833 TTTGCCCGGTTGCTCTTCTTTCTATCTTCCTTCGGCTTTGGCTTTGTTGACCATTCC   892

Qy   540 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCATGTCACCATGTCACGAATGCTGC   599
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   893 AGTTCCGGCTTATGAAGTGCGCAACGTGTCCATGTCATATACCATGTCACGACTGCTC   952

Qy   600 CAACTCAAGCATCGTTGGTTACGAGAGCACGAGATCATTGCACACCCCGGGGTGTGTGCC   659
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   953 CAACTCAAGCATCTGTTGTTACGAGAGCACGAGATCATTGCACACCCCGGGGTGTGTGCC  1012

Qy   660 CTGTGTCCCGGGAGGGGTAATTCCTCCCGGCTGCTCGGTAGGCGCTCACTCCACGCGCGGC   719
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1013 CTGCGGTTCGGGGGAGCAACTCCCGTCCCGGTTGCTCGGGTAGGCGCTCAGCGTCGCGC  1072

Qy   720 CAAGGACGGCCAGCAATACCCACCACTCGGACCGCCACGGTCGATTTGCTCGTTGGGGGC   779
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1073 CAGGAATGCCACGAATCCCCACTACGACACGACGGTCGGTTCCTCGTTGGGGGAC  1132

Qy   780 GGCTGCGCTTCTCTGTCCCGGCTGTATGTAGGGGGGATCTCTGCGGGATCTGTTTTCTCC   832
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1133 GGCTGCGCTTCTGTGCCCGCTATGTAGGTGGGGGGATCCTCGGGGGATCTGTTTTCC  1185
```

RESULT 15

```
ORIGIN
Query Match     89.3%;  Score 745;  DB 14;  Length 9379;
Best Local Similarity   95.0%;  Pred. No. 2.8e-171;
Matches 791;  Conservative   0;  Mismatches 40;  Indels  2;  Gaps  2;
```

```
RSVWKDLLEDTEPIDTTVMAKNEVFCIQPEKGGRKPARLIVFPDLGVRVCEKMALYD
VVSTLPQAVMGSSYGFQYSPSQRVEFLVNAWKSKKCPMGFSYDTRCFPDSTVTENDIRV
EESIYQCCDLAPEARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
TCYLKASAACRAAKLQDCTMLVNGDDLVVICESAGTQEDAANLRVFTEAMTRYSAPPG
DPPQPAYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
GNIIMYAPTLWARMVLMTHFFSIILAQEQLEKALDQIYGAHYSIEPLDLPQIIERLH
GLSAFSLHSYSPGEINRVAACLRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYL
FNWAVRTKLKTLPIPAASQDLLSSWVAGYSGGDIIHSLSRARPRWFMLCLLLLSVGV
GIYLLPNR"
```

```
Qy    1 AGAAAAACCAAACGTAACCAACCTCCGCCCCACCAGGAGCAGCAGCAGTCAGGTTCCGGGGGGCGGCGTGGT   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 AGAAAAACCAAACGTAACCAACCTCCGCCCCACCAGGAGCAGCAGCAGTCAGGTTCCGGGGGGCGGCGTGGT  413

Qy   61 CAGATCGTTGGTGGTGGAGTTTACCTTGCCGCGGCGCGCCCAGGTGGGTGTGCGCGCCG                120
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  414 CAGATCGTTGGTGGTGGAGTTTACCTTGCCGCGGCGCGCCCAGGTGGGTGTGCGCGCCG               473

Qy  121 ACTAGGAAGAGTTCCGAGCGGTCGCAACCTCGTTGGGAAGGGGACGAACCTTATTCCCCAAGGCT         180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  474 ACTAGGAAGAGTTCCGAGCGGTCGCAACCTCGTTGGGAAGGCGACAACCTTATTCCCCAAGGCT         533

Qy  181 CGCCAGCCCAGGGACCCCGGCGAGGGCAGGGCCCTGGGCTTCAGCCACCTCGTTGGCCCCTCTATGGCAAC  240
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  534 CGCCAGCCCAGGGGACCCCGGCGAGGGCAGGGCCCTGGGCTTCAGCCACCTCGTTGGCCCCTCTATGGCAAC 593

Qy  241 GAGGGCATGGGGGGTGGGGCAGGAGGATGGCTTCCTGTCTCACCGCGTTAGGTCATTGGGCAGC          300
        |||||||||||||||||||||||||||||||||||||||||||||
Db  594 GAGGGCATGGGGGGTGGGGCAGGAGGTTGGCTTCCTGTCTCATCGGAGTACCCTCACATGC           653

Qy  301 CCCACTGACCCCCGGCCTAGGTGTCGCGGTAAATTTGGGTAAAGTCATCATCGATTACCCTCACATGC     360
        |||||||||||||||||||||||||||||||||||||||||||||
Db  654 CCCACTGACCCCCGGCCGTAGGTGTCGCGGTAAATTTGGGTAAGGTCATCATCGATTACCCTCACATGC    713

Qy  361 GGCTTCGCGGACTCTCATGGGGGGGTACAATTCCCTGCGTCCTCCCCTTA-GGGGCGCTGC           419
        |||||||||||||||||||||||||||||||||||||| ||||||
Db  714 GGCTTCGCGGAC-CTCATGGGGGTACATTCCGCGTGCTCGGCGCTGC                          772

Qy  420 CAGGGCCCTGGGCCATGGCGTCCGGGGTTCTGGAGGAGACGGGTGAACATACAGAGGGAA           479
        |||||||||||||||||||||||||||||||||||||||||||||
Db  773 CAGGGCCCTGGGCACATGGTGTCCGGGGTTCTGGAGGAGCGGTGAAACTATGCAACAGAGGGAA        832

Qy  480 TTTACCCGGGTTGCTCTTTCTCTCATCTTCCTCTTGGCCTTTGCTGTCTGTTTGACCATTCC          539
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  833 TTTGCCCGGGTTGCTCTTTCTCTATCTTCCTCTTGGCCTTCGTTGACCATCCC                    892

Qy  540 AGCTTCCGCCTTATGAAGTGCGAACGTGTCGCGGGATCTACACAGTCAGGAACGAATTGCTC          599
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  893 AGCTTCCGCCTTATGAAGTGCGCAAGCGTGTCCCGGGATTATACCATGTCACGGAACGAACTGCTC       952

Qy  600 CAACTCAAGCATCGTGTACGACAGCGGCACATGATCATGCACACCCGGGTGTGTCC               659
        |||||||||||||||||||||||||||||||||||||||||||||
Db  953 CAACTCAAGCATTGTGTATGAGGCAGCGGCACTGATCATGCACACCCCGGGTGCGTCC            1012

Qy  660 CTGTGTTCCGGGGAGGGGTAATTCTTCCCCGCTGTCCTGCTCACTCCGCGCTCGCGGGGC           719
        |||||||||||||||||||||||||||||||||||||||||||||
Db 1013 CTGCCGTTGGGGAGAGCACTCCTCCCCCGTTGTCGTCACTCCACCCGCGCTCGCGGGGC           1072

Qy  720 CAAGGACACCCCAGCATCCCCACTCCGCACATACGAGCCGACAGTCCGGGCGAATTGCTCCGTTGGGGC      779
        |||||||||||||||||||||||||||||||||||||||||||||
Db 1073 CAGGAACGGCCCCAGCGTCCCCACTCACGACGCACAGTGCGACTTGCTCCGTTGGGGAC             1132

Qy  780 GGCTGCCTTCTCTCCGCTTCGGGGTGGGGGGATTCTCTGCGGATTCCGGTCTTCC                 832
        |||||||||||||||||||||||||||||||||||||||||||||
Db 1133 GGCTGCCTTTCTCGCTTATGTACGGTGGGGGGGATTCTGCGGGGATCTGTTTTCC                 1185
```

```
Db      RAGIYRFVFTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVWQ
        DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
Qy      LIRLKPTLHGPTPLLYRLGAVQNEVLTHPFTKYIMACMSADLEVVTSTWVLVGGVLA
        ALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYRREFDEMEECASHLPYIEQMQLAE
Db      QFKQKALGLLQTATKQAEAAAPVVESKWQALETFWAKHMWNFISGIQYLAGLSTLPGN
        PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPRAASAFVGAGIAGAAVGSI
Qy      GLGKVLVDILAGYGAGVAGALVAFKVMSGDMPSTEDLVNLLPAILSPGALVVGVVCAA
        ILRRHVGPGSGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
Db      RLHQWINEDCSTPCSGSWLRDVWDWICTVLSDFKTWLQSRVLPRLPGVPFLSCQRGYK
        GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSP
Qy      APNVSRALWRVAAEYVEVTRVGDFHYVTGMTTDNLKCPCQVPAPEFFTEALDGVRLHR
        YAPASKPLLRDEVTFQVGLNQYVVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRL
Db      ARGSPPSLASSSASQLSAPSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESEN
        KVVILDSFEPLRAEEDEREVSLPAEILRKSRKFPRAMPIWARPDYNPPLIESWKDPDY
Qy      VPPVVHGCPLPPTKAPPIPPPRRKRTVVLTESTVSSALAELATKTFGSSESSSAADSGT
        ATAPPDQPSDSGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMS
Db      YTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
        DHYRDVLKEMKAKASTVKARLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSRAINHI
Qy      RSVWKDLLQEDTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
        VVSTLPQAVMGAASYGYQYSPGQRVEFLVNAWKSKKCPMGFAYDTRCFDSTVTENDIRV
Db      EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
        TCYLKATAACRAAKLQDCTMLVCGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPG
Qy      DPPQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
        GNIIMFAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
Db      GLSAFSLHSYSPGEINRVAASCLRKLGVPPLRVWRHRARSRVRAKLLSQGGRAAICGKYL
        FNWAVRTKLKLTPIPAASQLDLSGWFVAGYSGGDIYHSVSRARPRWFMWCLLLLSVGV
        GIYLLPNR"
mat_peptide   330. .902
              /product="core protein"
mat_peptide   903. .1478
              /product="E1"
mat_peptide   1479. .2756
              /product="E2"
mat_peptide   2757. .3407
              /product="NS2"
mat_peptide   3408. .5300
              /product="NS3"
mat_peptide   5301. .5462
              /product="NS4A"
mat_peptide   5463. .6245
              /product="NS4B"
mat_peptide   6246. .7586
              /product="NS5A"
mat_peptide   7587. .9359
              /product="NS5B"
misc_feature  1479. .1559
              /note="hypervariable region (HVR)"
misc_feature  6954. .7073
              /note="interferon sensitivity determining region (ISDR)"

ORIGIN

Query Match      89.3%;   Score 745;  DB 14;  Length 9410;
Best Local Similarity  95.0%;  Pred. No. 2.8e-171;
Matches 791; Conservative  0; Mismatches 40; Indels  2; Gaps  2;

Qy        1 AGAAAACCAAACGTAACCAACCTCCCCCCACAGAGAGTCAGTTCCCGGGCGGGTTGGT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      354 AGAAAAACCAAACGTAACCAACCGCCCGCCCCCACAGAGACGTCAAGTTCCCGGGCGGGTTGGT 413

Qy       61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGTTGGGTGTGCGCGCG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      414 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGTTGGGTGTGCGCGCG 473

Qy      121 ACTAGGAAGAACTTCCGACGCGGTGCAACCTCGTTGGAAGCGACAAACCTATCCCCAAGCCT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      474 ACTAGGAAGAACTTCCGACGCGGTCCCAACCTCGTTGGAAGCGACAACCTATCCCCAAGGCT 533

Qy      181 CGCCAGCCCGAGGGGCCAGAGGGCCAGGGCTACCCCTTGGCCCTTCTATAGGCAAC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      534 CGCCAGCCCGAGGGGCAGGGGCCTGGGCTCAGCCCGGGTACCCCTTGGCCCTCTATAGGCAAT 593

Qy      241 GAGGGCATGGGGGTGGGGCAGGAGGATGGCTCTCCTGTCACCCCGTGGCTTGGGGGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      594 GAGGGCATGGGGGTGGGGCAGGAGGATGGCTCCTCCTGTCACCCCGCGGTCTCGGGGGCTCAGTTGGGGGC 653

Db      654 CCCACGGACCCCCGCGGGTAGGTCCGTGAATTGGGTAAGGTCATCGATCACATGC 713

Qy      361 GGCTTCGCCGGACTCTTCATGGGGGTACATTCCGCGTCGTCGTGCCCCCTTA-GGGGGCTGC 419
            |||| | |||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db      714 GGGTTCGCCGGAC-CTCACGGGGGGTACATTCCGCGTCGTCGTGCCCCCTTA-GGGGGCTTTC 772

Qy      420 CAGGGGCCCTGGGCGCATGGGCGTCCGGGTTCTTGGAGGACGCGGTGAACTATCGCAACAGGGAA 479
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      773 CAGGGGCCCTGGGCACATGGGCGTCCGGGGTTCTTGGAGGACGGGCGTGAACTATCGCAACAGGGAA 832

Qy      480 TTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCGTCGTTTGACCATTCC 539
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      833 TCTGCCCGGTTGCTCTTTCTCTATCTTCCTGGCTTTGCTGTCGTCGTTTGACCATCCC 892

Qy      540 AGCTTCCGCGCTTATGAAGTGCGCAACGTGTCCGGGGGATCTACCCATGTCACGGAACGATTGCTC 599
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      893 AGCTTCCGCGCTTATGAGGTGCGCAACGTGTCCGGGGGTACCATGTCACCATGTCACGGAACGACTGCTC 952

Qy      600 CAACTCAAGCGATCGTGTACGGAGACAGCGACATGATCATGCCACACCCCGGGGTGTGTGCC 659
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      953 TAACGCAAGCATTCGTGTATGAGGCAGCGGACATGATCATGCCACATCCCCCGGGGTGTGTGCC 1012

Qy      660 CTGTGTCCGGGCGAGGGTAATTCCTCCCGCTGCTCGGGGTAGCCGCTCACTCCCACGCTCGCGGC 719
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1013 CTGCGTTCCGGGAGGGCCAATTCCTCCCCGCTGCTCGGGGTAGCCGCTCACTCCCACGCTCGCGGC 1072

Qy      720 CAAGGACGCCAGCANTCCCCACTGCGACAATACGACGCCCAGGTCGATTTGCTCGTTGGGGG 779
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1073 CAGGAACTCCAGCGTCCCAACTACGACAATACGACGCCCAGGTCGATTTGCTCGTTGGGGGC 1132

Qy      780 GGCTGCCTTCTTCGTCCCGCTATGTACGTGGGGGGATCTCTCGGGATCTGTTTTTTCC 832
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1133 GGCTGCTTTCTGCTCCCGCTATGTACGTGGGGGGATCTCTCGGGATCTGTTTTTTCC 1185
```

Search completed: February 21, 2005, 03:06:16
Job time : 4119 secs

us-09-664-363-5.rge

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: February 20, 2005, 21:58:35 ; Search time 588 Seconds
                   (without alignments)
                   8396.368 Million cell updates/sec

Title: US-09-664-363-5
Perfect score: 834
Sequence: 1 AGAAAAACCAAAACGTAACAC..........TCTCGGGATCTGTTTTCCCG 834

Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0
Searched: 4390206 seqs, 2559870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Database :   N Geneseq_16Dec04:*
                   1: geneseqn1980s:*
                   2: geneseqn1990s:*
                   3: geneseqn2000s:*
                   4: geneseqn2001as:*
                   5: geneseqn2001bs:*
                   6: geneseqn2002as:*
                   7: geneseqn2002bs:*
                   8: geneseqn2003as:*
                   9: geneseqn2003bs:*
                   10: geneseqn2003cs:*
                   11: geneseqn2003ds:*
                   12: geneseqn2004as:*
                   13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	834	100.0	834	2	AAQ12238	Aaq12238 Clone BR1
2	804.8	96.5	2790	2	AAQ46191	Aaq46191 PT-NANBH
3	804.8	96.5	3372	2	AAQ46202	Aaq46202 PT-NANBH
4	801.6	96.1	831	2	AAQ46192	Aaq46192 PT-NANBH
5	799.4	95.9	2116	2	AAQ12242	Aaq12242 Encodes P
6	738.6	88.6	1734	2	AAQ40428	Aaq40428 Hepatitis
7	738.6	88.6	1734	2	AAQ40429	Aaq40429 Hepatitis
8	737	88.4	1734	2	AAQ40430	Aaq40430 Hepatitis
9	737	88.4	1734	2	AAQ40438	Aaq40438 Hepatitis
10	737	88.4	2433	2	AAT12974	Aat12974 HCV E1 co
11	737	88.4	2433	10	ADS55557	Ads55557 Hepatitis
12	737	88.4	2433	12	ADP71139	Adp71139 HCV cDNA
13	737	88.4	9587	13	ADR82189	Adr82189 Hepatitis
14	733.8	88.0	2187	2	ABA03491	Aba03491 Cuticle P
15	733.8	88.0	2540	2	AAQ43889	Aaq43889 NANB hepa
16	733.8	88.0	3360	2	AAT03677	Aat03677 Hepatitis
17	733.8	88.0	3461	2	AAQ64068	Aaq64068 Non-A, no
18	733.8	88.0	3461	2	AAT30386	Aat30386 5'UTR/COR
19	733.8	88.0	9413	2	AAQ81559	Aaq81559 Hepatitis
20	733.8	88.0	9413	2	AAT03960	Aat03960 Partial H
21	733.8	88.0	9413	6	AAD25517	Aad25517 Hepatitis
22	733.8	88.0	9413	8	AAL53723	Aal53723 Hepatitis
23	733.8	88.0	9413	8	AAD49655	Aad49655 Hepatitis
24	733.8	88.0	9413	10	ADF88596	Adf88596 Hepatitis
25	733.8	88.0	9611	13	ADS34713	Ads34713 Hepatitis
26	732.2	87.8	1880	2	AAQ24467	Aaq24467 NANB hepa
27	732.2	87.8	2540	2	AAQ63753	Aaq63753 NANBH ge
28	732.2	87.8	3401	2	AAQ64069	Aaq64069 Non-A, no
29	733.8	88.0	9611	13	AAT30387	Aat30387 5'UTR/COR
30	730.6	87.6	1270	2	AAV60668	Aav60668 Fragment
31	730.6	87.6	1562	2	AAV60672	Aav60672 Fragment
32	730.6	87.6	1734	2	AAQ40437	Aaq40437 Hepatitis
33	730.6	87.6	1953	8	AAL53222	Aal53222 Plasmid p
34	730.6	87.6	2829	2	AAQ29628	Aaq29628 Hepatitis
35	730.6	87.6	2540	2	AAV60673	Aav60673 Fragment
36	730.6	87.6	9413	2	AAQ80498	Aaq80498 DNA encod
37	730.6	87.6	9595	4	AAF23492	Aaf23492 Infection
38	730.6	87.6	9595	2	AAX24843	Aax24843 Infectiou
39	730.6	87.6	9595	12	AAC86939	Aac86939 Nucleotid
40	730.6	87.6	9595	12	ADO36222	Ado36222 Hepatitis
41	730.6	87.6	9599	2	ADO79396	Ado79396 Hepatitis
42	730.6	87.6	9605	2	AAX24833	Aax24833 Infectiou
43	730.6	87.6	9605	6	ABK91431	Abk91431 Hepatitis
44	730.6	87.6	9605	6	ABK91424	Abk91424 Hepatitis
45	730.6	87.6	9605	6	ABK91429	Abk91429 Hepatitis

ALIGNMENTS

RESULT 1
AAQ12238
ID   AAQ12238 standard; DNA; 834 BP.
XX
AC   AAQ12238;
XX
DT   25-MAR-2003 (revised)
DT   06-SEP-1991 (first entry)
XX
DE   Clone BR11 encoding PT-NANBH virus antigenic portion.
XX
KW   post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.
XX
OS   Non-A, non-B hepatitis virus.
XX
XX
PN   GB2239245-A.
XX
PD   26-JUN-1991.
XX
PF   17-DEC-1990; 90GB-00027250.
XX
PR   18-DEC-1989; 89GB-00028562.
PR   27-FEB-1990; 90GB-00004414.
PR   03-MAR-1990; 90GB-00004814.
XX
PA   (WELL) WELLCOME FOUND LTD.
PA   (HIGH/) HIGHFIELD P E.
XX
PI   Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ;
XX
DR   WPI; 1991-187584/26.
DR   P-PSDB; AAR12596.
XX
PT   Post-transfusional non-A, non-B hepatitis poly:peptide(e) - and also
PT   and antibodies used in diagnostic assays and in vaccines.
XX
PS   Claim 10; Page 56-58; 108pp; English.
XX
CC   This sequence is a structural region of the PT-NANBH viral genome
CC   encoding an antigenic polypeptide. It was isolated from serum A cDNA
CC   library was prepared in lambda gt11 from the serum of infected patients
CC   and screened with antibodies from the serum of humans with a high risk

```
CC   for PT-NANBH but which did not react with viral antigens DX113, BHC-5 and
CC   BHC-7. Clone BR11 was identified which did not cross-hybridise with
CC   probes made from JG2 and JG3 (see AAQ12236 and AAQ12237). It was
CC   sequenced and found to have the sequence shown, which includes the EcoRI
CC   linkers added during cloning. See also AAQ12239-Q12242. (Updated on 25-
CC   MAR-2003 to correct PA field.)
XX
SQ   Sequence 834 BP; 139 A; 265 C; 251 G; 179 T; 0 U; 0 Other;

Query Match             100.0%; Score 834; DB 2; Length 834;
Best Local Similarity   100.0%;
Matches   834; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy          1 AGAAAAACCAAACGTAACACCAAACCTCCGCCACCAGAACGTCAGGTTCCCGGCGGTCGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAAAAACCAAACGTAACACCAAACCTCCGCCACCAGAACGTCAGGTTCCCGGCGGTCGT 60

Qy         61 CAGATCGTTGGTGTGGAGTTTACCTGTTGCCGCGGCGCCCCAGGTTGGGTGTGCGGCGCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGATCGTTGGTGTGGAGTTTACCTGTTGCCGCGGCGCCCCAGGTTGGGTGTGCGGCGCG 120

Qy        121 ACTAGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGTGGAAGGGACGAACCTTATCCCAAGGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTAGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGTGGAAGGGACGAACCTTATCCCAAGGCT 180

Qy        181 CGCCAGCCCGAGGGCAGGGCGTTGGGCTCAGCCCTTGGGCCCCTCTATGGCAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCCAGCCCGAGGGCAGGGCGTTGGGCTCAGCCCTTGGGCCCCTCTATGGCAAC 240

Qy        241 GAGGGCCATGGGGTGGGCCAGGAGGATGGGTCCTCGTGTCCACCCGTGTCCCGGCCTAGTTGGGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGGGCCATGGGGTGGGCCAGGAGGATGGGTCCTCGTGTCCACCCGTGTCCCGGCCTAGTTGGGGC 300

Qy        301 CCCACTGACCCCCCGGGGGTAGGTCGGGTAAATTGGGGTAAAGTCATGGATACCCTCACATGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCACTGACCCCCCGGGGGTAGGTCGGGTAAATTGGGGTAAAGTCATGGATACCCTCACATGC 360

Qy        361 GGCCTTCGCCGACTCTCATGGGGGTTACATTCCGGCTCGTGCGGCGGCCTTCGCCTGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCCTTCGCCGACTCTCATGGGGGTTACATTCCGGCTCGTGCGGCGGCCTTCGCCTGCC 420

Qy        421 AGGGCCCTGGGCGGTCGGGTTCGGAGGACACGGCGTGAACTATGCAACAGGGAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGGCCCTGGGCGGTCGGGTTCGGAGGACACGGCGTGAACTATGCAACAGGGAAT 480

Qy        481 TTACCCGGTTGGCTTCTCTATCCTCCTTGGCTTGGTCGCGGTCCTGTTTGACCATTCCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTACCCGGTTGGCTTCTCTATCCTCCTTGGCTTGGTCGCGGTCCTGTTTGACCATTCCA 540

Qy        541 GCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGGATCTTCCATGTCACGAACGATTGCTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGGATCTTCCATGTCACGAACGATTGCTCC 600

Qy        601 AACTCAAGCATCGTGTTACGAGACAGCGGCACATGATCATGCACCCCCGGGGTGGTCGTGCCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTCAAGCATCGTGTTACGAGACAGCGGCACATGATCATGCACCCCCGGGGTGGTCGTGCCC 660

Qy        661 TGTGTCCGGGAGGGGTAATTCCTCCCGCTGCTCGGGTAGCGCTCACTCCCAGCGTCCGCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGTGTCCGGGAGGGGTAATTCCTCCCGCTGCTCGGGTAGCGCTCACTCCCAGCGTCCGCGCC 720

Qy        721 AAGGACGCCCAGCATCCCACTGCGACAATACGACGCACGTCGATTTGCTCGTTCGTTGGGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAGGACGCCCAGCATCCCACTGCGACAATACGACGCACGTCGATTTGCTCGTTCGTTGGGGCG 780

Qy        781 GCTGCCTTCTCTCGCCGCTATGTACGTGGGGGGATCTCTCGCGGATCTGTTTTCCCG 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTGCCTTCTCTCGCCGCTATGTACGTGGGGGGATCTCTCGCGGATCTGTTTTCCCG 834
```

```
AAQ46191
ID   AAQ46191 standard; DNA; 2790 BP.
XX
AC   AAQ46191;
XX
DT   27-AUG-2003 (revised)
DT   25-MAR-2003 (revised)
DT   24-FEB-1994 (first entry)
XX
DE   PT-NANBH virus BHC-11 fusion protein.
XX
KW   Parenterally transmitted non A non B hepatitis; PT-NANBH;
KW   hepatitis C virus; HCV; NS5; E1; linker; detection; diagnosis; antigen;
KW   vaccine; BHC-11; replicase; core protein;
KW   Autographa californica nuclear polyhedrosis virus; AcNPV; polyhedrin; ss.
XX
OS   Hepatitis virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2790
FT                   /tag= a
FT                   /product= "fusion protein"
FT   misc_feature    1..63
FT                   /tag= b
FT                   /label= AcNPV_polyhedrin_N-terminal
FT   misc_feature    64..1851
FT                   /tag= c
FT                   /label= PT-NANBH_NS5
FT   misc_feature    1852..1875
FT                   /tag= d
FT                   /note= "Synthetic linker region"
FT   misc_feature    1876..2706
FT                   /tag= e
FT                   /label= PT-NANBH_core_and_E1_regions
FT   misc_feature    2707..2790
FT                   /tag= f
FT                   /note= "Polyhedrin gene sequence read out-of-frame"
XX
PN   WO9317110-A2.
XX
PD   02-SEP-1993.
XX
PF   19-FEB-1993; 93WO-GB000345.
XX
PR   21-FEB-1992; 92GB-00003803.
XX
PA   (WELL ) WELLCOME FOUND LTD.
XX
PI   Parker D, Rodgers BC;
XX
DR   WPI; 1993-288415/36.
DR   P-PSDB; AAR41431.
XX
PT   New recombinant polyepptide for diagnosing hepatitis C - contains three
PT   distinct antigens from different viral regions, also useful in protective
PT   vaccines.
XX
PS   Disclosure; Page 20-24; 99pp; English.
XX
CC   GB-A-2239245 discloses a recombinant polypeptide BHC-11 which comprises
CC   an antigen obtained from the non-structural coding region (NS) (the 3'
CC   end) and one antigen from the structural coding region (S) (the 5' end)
CC   of the NANBH virus. Specifically BHC-11 (AAQ46191) contains a portion of
CC   the non-structural region of the virus, called NS5, (putative replicase)
CC   at the N-terminus joined via a synthetic linker to a portion of the
CC   structural region which contains almost all the core protein sequence (9
CC   amino acids from the N-terminal are not present) and a part of a sequence
CC   from the structural region called E1. It is disclosed that BCH-11 may be
CC   used in diagnosis of PT-NANBH. If at least three different PT-NANBH
CC   antigens are used to screen for PT-NANBH, the screening is much more
CC   sensitive as compared to the use of only two PT-NANBH antigens. Pref.
CC   antigens are described in AAQ46192-94. Two new antigenic regions of the
CC   PT-NANBH genome are given in AAQ46198-99. AAQ46202 describes an improved
CC   PT-NANBH recombinant polypeptide. (Updated on 27-AUG-2003 to correct PN
CC   field.) (Updated on 25-MAR-2003 to correct OS field.)
```

RESULT 3

```
ID   AAQ46202 standard; cDNA to mRNA; 3372 BP.
XX
AC   AAQ46202;
XX
DT   24-FEB-1994  (first entry)
DT   25-MAR-2003  (revised)
DT   27-AUG-2003  (revised)
XX
DE   PT-NANBH NS5-NS3-core recombinant polypeptide.
XX
KW   Parenterally transmitted non A non B hepatitis; PT-NANBH;
KW   hepatitis C virus; HCV; NS3; NS5; core; E1; linker; detection; diagnosis;
KW   antigen; vaccine; BHC-28; replicase; core protein;
KW   Autographa californica nuclear polyhedrosis virus; AcNPV; polyhedrin; ss.
XX
OS   Hepatitis virus.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..3372
FT                   /*tag= a
FT                   /product= "fusion_protein"
FT   misc_feature    1..63
FT                   /*tag= b
FT                   /label= AcNPV_polyhedrin_N-terminal
FT   misc_feature    64..1852
FT                   /*tag= c
FT                   /label= PT-NANBH_NS5
FT   misc_feature    1853..1858
FT                   /*tag= d
FT                   /note= "PstI restriction site"
FT   misc_feature    1859..2434
FT                   /*tag= e
FT                   /label= PT-NANBH_NS3
FT   misc_feature    2435..2457
FT                   /*tag= f
FT                   /note= "PsrI restriction site"
FT   misc_feature    2458..3288
FT                   /*tag= g
FT                   /label= PT-NANBH_core_and_E1_regions
FT   misc_feature    3289..3272
FT                   /*tag= h
FT                   /note= "Polyhedrin gene sequence read out-of-frame"
XX
PN   WO9117110-A2.
XX
PD   02-SEP-1993.
XX
PF   19-FEB-1993;   93WO-GB000345.
XX
PR   21-FEB-1992;   92GB-00003803.
XX
PA   (WELL ) WELLCOME FOUND LTD.
XX
PI   Parker D, Rodgers BC;
XX
DR   WPI; 1993-288415/36.
DR   P-PSDB; AAR41439.
XX
PT   New recombinant polypeptide for diagnosing hepatitis C - contains three
PT   distinct antigens from different viral regions, also useful in protective
PT   vaccines.
XX
PS   Claim 7; Page 85-90; 99pp; English.
XX
CC   The NS3 specific region of pDX200 was amplified by PCR using primers D360
CC   and D361 (AAQ46201 and AAQ54882). The obtained fragment was then cloned
CC   into pDX136. Transformants were analysed to identify those which
CC   contained the NS3 sequence inserted in the correct orientation between
CC   the NS5 and core parts of pDX136, this was called pDX208 (AAQ46202).
CC   Recombinant baculovirus BHC-28 was produced. Insect cells infected with
CC   BHC-28 produce antigen NS5-NS3-core. If at least three different PT-NANBH
CC   antigens are used to screen for FP-NANBH, the screening is much more
CC   sensitive as compared to the use of only two PT-NANBH antigens. Pref.
CC   antigens are described in AAQ46192-94. Two new antigenic regions of the
CC   PT-NANBH genome are given in AAQ46198-99. AAQ46202 describes an improved
CC   PT-NANBH recombinant polypeptide. (Updated on 25-MAR-2003 to correct PN
CC   field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 3372 BP; 714 A; 1026 C; 954 G; 678 T; 0 U; 0 Other;
```

```
XX
SQ   Sequence 2790 BP; 585 A; 861 C; 789 G; 555 T; 0 U; 0 Other;

Query Match      96.5%;  Score 804.8;  DB 2;  Length 2790;
Best Local Similarity   99.5%;  Pred. No. 4.8e-208;
Matches  828;  Conservative   0;  Mismatches   2;  Indels   2;  Gaps   2;

Qy      4 AAAACCAAACGTAACACCAACCTCCGCCCACCACAGAGACGTCAGGTTCCCGGACGGGTCAG 63
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1876 AAAACCAAACGTAACACCAACCTCCGCCCACCACAGAGACGTCAGGTTCCCGGACGGGTGGTCAG 1935

Qy     64 ATCGTTGGTGGGAGTTACCTGTTGCCGGCCCAGGTGGGTGGTGCGGCCGGACT 123
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1936 ATCGTTGGTGGGAGTTACCTGTTGCCGGCCCAGGTGGGTGGTGCGGCCGGACT 1995

Qy    124 AGGAAGACTTCCGAGCGGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAGGCTCGC 183
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1996 AGGAAGACTTCCGAGCGGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAGGCTCGC 2055

Qy    184 CAGCCCGGAGGGAAGGCGCAGGGCGTACCCTTGGCCCTCTATGCCAACGAG 243
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2056 CAGCCCGGAGGGAAGGCGCAGGGCGTACCCTTGGCCCTCTATGCCAACGAG 2115

Qy    244 GGCATGGGGGTGGGGCAGGATGGCTCCTGTCACCCCGGGCCCTAGTTGGGGCCCC 303
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2116 GGCATGGGGGTGGGGCAGGATGGCTCCTGTCACCCCGGGCCCTAGTTGGGGCCCC 2175

Qy    304 ACTGACCCCCGGGCCTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGCGGC 363
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2176 ACTGACCCCCGGGCCTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGCGGC 2235

Qy    364 TTCGCCCGACTCTCATGGGGGTACAATTCCGCTCGTCGTCCCTTA-GGGGCGTGCCAG 422
          |||||||||| |||||||||||||||||||||||||||         |||||||||||
Db   2236 TTCGCCCGAC-CTCATGGGGTACAATTCCGCTCGTCGTGGGGGCGTGCCAG 2294

Qy    423 GGGCCCTGGGGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGAATTT 482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2295 GGGCCCTGGGGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGAAATTT 2354

Qy    483 ACCCGGTTGCTCTTTCTCTTATCTTCCTCTTGCTTTGTCGTTTGACCCATTCCAGC 542
          ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||
Db   2355 ACCCGGTTGCTCTTTCTCTTATCTTCCTCTTGCTTTGTCGTCTCTGTTTGACCCATTCCAGC 2414

Qy    543 TTCCGGCTTATGAAGTGCGGCAACGTGTCTCCGGGAATCTACCATGTCACGAACGGATTGCTCCAA 602
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   2415 TTCCGGCTTATGAAGTGCGGCAACGTGTCTCCGGGGATCTACCATGTCACGAACGGATTGCTCCAA 2474

Qy    603 CTCAAGACATCGTGTACGAGACAGCGGACAATGACTCATGCACACCCCGGTGTGTTGCCCTG 662
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2475 CTCAAGACATCGTGTACGAGACAGCGGACAATGACTCATGCACACCCCGGTGTGTTGCCCTG 2534

Qy    663 TGTCCCGGAGGGTAATTCCTCCCGCCTGTGGGGTAGCGCTCACTCCACCTCGCGGCCAA 722
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2535 TGTCCCGGAGGGTAATTCCTCCCGCCTGTGGGGTAGCGCTCACTCCACCTCGCGGCCAA 2594

Qy    723 GGAGCGCCAGCATCCCACTGCGACAATACGACGCCACGTCGATTGCTGGTTGGGGCGGC 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2595 GGAGCGCCAGCATCCCACTGCGACAATACGACGCCACGTCGATTGCTGGTTGGGGCGGC 2654

Qy    783 TGCCTTCTCGTCCCTATGTACGTGGGGGGATCTCTGCGGATCTGTTTTCCG 834
          |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   2655 TGCCTTCTGCTCCCTATGTACGTGGGGGGATCTCTGCGGATCTGTTTTCCG 2706
```

Query Match   96.5%;   Score 804.8;   DB 2;   Length 3372;
Best Local Similarity   99.5%;   Pred. No. 5.1e-208;
Matches 828;   Conservative   0;   Mismatches   2;   Indels   2;   Gaps   2;

```
Qy      4 AAAACCAAAACGTAACACCAAACCTCCGCCCACCAGGACGTCAGGTTCCCGGCGTGGGTCAG 63
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2458 AAAACCAAAACGTAACACCAAACCTCCGCCCACCAGGACGTCAGGTTCCCGGCGTGGGTCAG 2517

Qy     64 ATCGTTGGTGGAGTTTACCTGTTGCCCGGCGACAGGGGCCCCCAGGGTTGGGTGTGCCCGACT 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2518 ATCGTTGGTGGAGTTTACCTGTTGCCCGGCGACAGGGGCCCCCAGGGTTGGGTGTGCCCGACT 2577

Qy    124 AGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGGCGACAACCTATCCCCAAGGCTTCGC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2578 AGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGGCGACAACCTATCCCCAAGGCTTCGC 2637

Qy    184 CAGCCCCGAGGGCAGGGGCCTGGGCCTCAGCCCCGGGTACCCTTCAGCCCCTCTATGGCAACGAG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2638 CAGCCCCGAGGGCAGGGGCCTGGGCCTCAGCCCCGGGTACCCTTCAGCCCCTCTATGGCAACGAG 2697

Qy    244 GGCCATGGGGGTGGGCAGAGATGGCTCCCGTGCTCCCGGCGTTCTCGGAGGACGCGGGAATTT 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2698 GGCCATGGGGGTGGGCAGAGATGGCTCCCGTGCTCCCGGCGTTCTCGGAGGACGCGGGAATTT 2757

Qy    304 ACTGACCCCCGGGGGTAGGTCGGGTAATTTGGGGTAAAGTCATCGGATAACCCTCACATGCGGC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2758 ACTGACCCCCGGGGGTAGGTCGGGTAATTTGGGGTAAAGTCATCGGATAACCCTCACATGCGGC 2817

Qy    364 TTCGCCGGACTCTCATGGGGTACATTCCCGCTCGTCGGCGCCTCCCTTA-CGGGGCCGCCTGCCAG 422
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   2818 TTCGCCGGAC-CTCATGGGGTACATTCCCGCTCGTCGGCGCCTCCCTTAGGGGGCCGCCTGCCAG 2876

Qy    423 GGCCCCTGGCGCAATGGCGTCCGGGTTCTGGAGGACGCGGGAATTTCGAACAGGGAATTT 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2877 GGCCCCTGGCGCAATGGCGTCCGGGTTCTGGAGGACGCGGGAATTTCGAACAGGGAATTT 2936

Qy    483 ACCGGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTGCTTGTCGTTTTGACCATTCCAGC 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2937 ACCGGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTGCTTGTCGTTTTGACCATTCCAGC 2996

Qy    543 TTCCGCTTATGAAGTGCCCAACCTGTCCGGGATCTACCATGTCACGAACGAATTGCTCCAA 602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2997 TTCCGCTTATGAAGTGCCCAACCTGTCCGGGATCTACCATGTCACGAACGAATTGCTCCAA 3056

Qy    603 CTCAAGCAATCGTGTACGAGAGACAGCGGGACATGACATGCACACCCCGGGTGTCTGCCCTG 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3057 CTCAAGCAATCGTGTACGAGAGACAGCGGGACATGACATGCACACCCCGGGTGTCTGCCCTG 3116

Qy    663 TGTCCGGGGAGGGGTTAATTCCTCCCGGGGGTTAGCGCTCACTCCCAGCTCGCGGGCCAA 722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3117 TGTCCGGGGAGGGGTTAATTCCTCCCGGGGGTTAGCGCTCACTCCCAGCTCGCGGGCCAA 3176

Qy    723 GGACGCGCAGCATCCCACTCCCACGTCGATTTGCTCGTTGGGGCGGCC 782
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3177 GGACGCGCAGCATCCCACTCCCACGTCGATTTGCTCGTTGGGGCGGCC 3236

Qy    783 TGCCTTCTGCTCCGCTAGTAGTAGGGGGATCTCTGCGGATCTGTTTTTCCCG 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3237 TGCCTTCTGCTCCGCTAGTAGTAGGGGGATCTCTGCGGATCTGTTTTTCCCG 3288
```

Query Match   96.1%;   Score 801.6;   DB 2;   Length 831;
Best Local Similarity   99.3%;   Pred. No. 2.4e-207;
Matches 826;   Conservative   0;   Mismatches   4;   Indels   2;   Gaps   2;

```
Qy      4 AAAACCAAAACGTAACACCAAACCTCCGCCCACCAGGACGTCAGGTTCCCGGCGTGGGTCAG 63
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAACCAAAACGTAACACCAAACCTCCGCCCACCAGGACGTCAGGTTCCCGGCGTGGGTCAG 60

Qy     64 ATCGTTGGTGGAGTTTACCTGTTGCCCGGCGACAGGGGCCCCCAGGGTTGGGTGTGCCCGACT 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCGTTGGTGGAGTTTACCTGTTGCCCGGCGACAGGGGCCCCCAGGGTTGGGTGTGCCCGACT 120

Qy    124 AGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGGCGACAACCTATCCCCAAGGCTTCGC 183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGAAGAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGGCGACAACCTATCCCCAAGGCTTCGC 180

Qy    184 CAGCCCCGAGGGCAGGGGCCTGGGCCTCAGCCCCGGGTACCCTTCAGCCCCTCTATGGCAACGAG 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGCCCCGAGGGCAGGGGCCTGGGCCTCAGCCCCGGGTACCCTTCAGCCCCTCTATGGCAACGAG 240

Qy    244 GGCCATGGGGGTGGGCAGAGATGGCTCCCGTGCTCCCGGCGTTCTCGGAGGACGCGGGAATTT 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCCATGGGGGTGGGCAGAGATGGCTCCCGTGCTCCCGGCGTTCTCGGAGGACGCGGGAATTT 300

Qy    304 ACTGACCCCCGGGGGTAGGTCGGGTAATTTGGGGTAAAGTCATCGGATAACCCTCACATGCGGC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACTGACCCCCGGGGGTAGGTCGGGTAATTTGGGGTAAAGTCATCGGATAACCCTCACATGCGGC 360

Qy    364 TTCGCCGGACTCTCATGGGGTACATTCCCGCTCGTCGGCGCCTCCCTTA-CGGGGCCGCCTGCCAG 422
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    361 TTCGCCGGAC-CTCATGGGGTACATTCCCGCTCGTCGGCGCCTCCCTTAGGGGGCCGCCTGCCAG 419

Qy    423 GGCCCCTGGCGCAATGGCGTCCGGGTTCTGGAGGACGCGGGAATTT 482
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    420 GGCCCCTGGCGCAATGGCGTCCGGGTTCTGGAGGACGCGGGAATTT 479

Qy    483 ACCGGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTGCTTGTCGTTTTGACCATTCCAGC 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 ACCGGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTGCTTGTCGTTTTGACCATTCCAGC 542
```

RESULT 5
AAQ12242
```
ID   AAQ12242 standard; DNA; 2116 BP.
XX
AC   AAQ12242;
XX
DT   25-MAR-2003 (revised)
DT   17-SEP-1991 (first entry).
XX
DE   Encodes PT-NANBH viral structural and non-structural proteins.
XX
KW   post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.
XX
OS   non-A.
OS   non-B hepatitis virus.
XX
FH   Key            Location/Qualifiers
FT   CDS            308..2116
FT                  /*tag= a
XX
CC   GB239245-A.
XX
PN
PD   26-JUN-1991.
XX
PF   17-DEC-1990;   90GB-00027250.
XX
PR   18-DEC-1989;   89GB-00028562.
PR   27-FEB-1990;   90GB-00004414.
PR   03-MAR-1990;   90GB-00004814.
XX
PA   (WELL ) WELLCOME FOUND LTD.
PA   (HIGH/) HIGHFIELD P E.
XX
PI   Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ;
XX
DR   WPI; 1991-187584/26.
DR   P-PSDB; AAR12600.
XX
PS   Claim 10; Page 83-87; 108pp; English.
XX
PT   Post-transfusional non-A non-B hepatitis poly:peptide(s) - and also DNA
PT   and antibodies used in diagnostic assays and in vaccines.
XX
CC   This sequence is thought to encode viral structural and non-structural
CC   proteins of the PT-NANBH viral genome which are antigenic. It was
CC   isolated from human serum infectious for the virus. See also AAQ12236-41.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 2116 BP; 392 A; 650 C; 624 G; 450 T; 0 U; 0 Other;
```

```
Query Match             95.9%;  Score 799.4;  DB 2;  Length 2116;
Best Local Similarity   99.0%;  Pred. No. 1.3e-206;
Matches  825;  Conservative  0;  Mismatches  6;  Indels  2;  Gaps  2;
```

(sequence alignment — Qy / Db paired lines with positional numbering, nucleotide sequences, match bars)

RESULT 6
AAQ40428
```
ID   AAQ40428 standard; cDNA; 1734 BP.
XX
AC   AAQ40428;
XX
DT   30-JUL-1993 (first entry)
XX
DE   Hepatitis C virus clone JK2-A.
XX
KW   HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
```

(sequence alignment — Qy / Db paired lines with positional numbering, nucleotide sequences, match bars)

KW   polymerase chain reaction; diagnostic method; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key   Location/Qualifiers
FT   CDS   325..1734
FT   /*tag= a
FT   /note= "not full-length CDS, hence no termination codon"
XX
PN   JP05068562-A.
XX
PD   23-MAR-1993.
XX
PF   30-MAY-1991;   91JP-00153736.
XX
PR   30-MAY-1991;   91JP-00153736.
XX
PA   (SANW) SANWA KAGAKU KENKYUSHO CO.
XX
CC   AAQ040425-Q40439
XX
CC   amplification, including clone JK2-A. Primer/probes derived from the
CC   nucleotides long) and 14 shorter clones were isolated by PCR
CC   cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B (9405
CC   sequences of these clones can be used in diagnostic assays for HCV. See
PS   Claim 4; Page 20-22; 44pp; Japanese.
PT   DNA and cDNA of hepatitis C virus - useful as probes for diagnosing HCV
PT   infection.
XX
DR   WPI; 1993-130638/16.
DR   P-PSDB; AAR34470.
XX
SQ   Sequence 1734 BP; 326 A; 537 C; 505 G; 366 T; 0 U; 0 Other;

Query Match   88.6%; Score 738.6; DB 2; Length 1734;
Best Local Similarity   94.5%; Pred. No. 3.9e-190;
Matches 787; Conservative 0; Mismatches 44; Indels 2; Gaps 2;

```
Qy        1 AGAAAAACCAAAAGTAACCAAACCTTCGCCCCACAGAGGAGTCAGGTTCCCGGCGGTGGT 60
            |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      349 AGAAAAACCAAAAGTAACCAAACCGCGCGCCACAGCGCAAGTTCCCGGCGGGTGGC 408

Qy       61 CAGATCGTTGGTGGAGTTTACCTGTTGCGCGCGCAGGGGGCCCCAGGGGGGTGCGCGCCG 120
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      409 CAGATCGTTGGTGGAGTTTACCTGTTGCGCGCGCAGGGGGCCCCAGGGTTGGGTGTGCGCGCCG 468

Qy      121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGGAAGGCGACAAACCTATCCCCAGGCCT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      469 ACTAGGAAGACTTCCGACGGTGCAACCTCGTGGAAGGCGACAAACCTATCCCCAGGCCT 528

Qy      181 CGCCAGCCCGAGGGCAGGGGCAGGGCCCTCAGGGTACCCTTGGGCCCCTCTATGGCAAC 240
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      529 CGCCAGCCCGAGGGCAGGGGCCCTGGGGCTCAGCGCCGGGTACCCTTGGGCCCCCTATGGCAAT 588

Qy      241 GAGGGCATGGGGTGGGCGCAGGATGGCTCTCGTTGTCACCCGGGCGCCTCAGTTGGGGC 300
            |||||||||||||||||||||||||||||||||||||||||||
Db      589 GAGGGCTCTGGGGTGGGCAGGGATGGCTCCTCGTCCCGCGCCTCAGTTGGGGGC 648

Qy      301 CCCACTGACCCCCCGCGTAGGTCGCGTTGGGAAAGTCATCGATACCCTTCACATGC 360
            ||||||||||||||||||||||||||||||||||||||||
Db      649 CCCACGACCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTCATCGATACCCTTCACATGC 708

Qy      361 GGCTTCGCCGACTCTCATGGGGTACATTCCGGTCGCGGCGGTTCCCTTAGG-GGCGGTGC 419
            ||||||||||||||||||||||||||||||||||||| |||||
Db      709 GGCTTCGCCGAC-CTCATGGGGTACATTCCGGTCGCGTCCGGTTCGGCGGGCTGC 767

Qy      420 CAGGGGCCCCTGGCGCCATGGCGTCCCGGTTCTGGAGGACCGTGAACATAGCAACAGGGAA 479
            |||||||||||||||||||||||||||||||||||||||||||||||
Db      768 CAGGGGCCCTGGGCGCCGCCATGGCGGTCCCGGTTCTGTGGAGGACGGCGTGAACTACGCAACAGGGAA 827

Qy      480 TTTACCCGGTTGCTCTTTTCTCTATCTTCCTCTCTGTCTCTGTTTGACCATTCC 539
            |||||||||||||||||||||
Db      828 TCTGCCCGGTTGCTCTTTTCTCTATCTTCCTCTCTGTTGGCCTCTTGTGTTCTGTGTTTGACCATTCC 887

Qy      540 AGCTTCCGGCTTATGGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGGAACGATTGCTC 599
            ||| ||||||||||||||||||||||||||||||||||
Db      888 AGCATCCGGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGGAACGACTGCTC 947

Qy      600 CAACTCAAGCCATCGTGTACGGACAGCGGCACATGATCATGCACACCCCCGGTGTGCC 659
            |||||||||||||||||||||||||||||||| |||||
Db      948 CAACTCAAGCTATATTGTGTATGTGAGAGACAGCGGAGCAGCGGACATTCATCATGCACACCCCCGGGTTGCC 1007

Qy      660 CTGTGTCCGGGGAGGGTAATTCCTCCCGCGTGGTCAGGTTAGGGCTTCACTCCCACGCTCC 719
            |||||||||||||||||||||||||||||||||
Db     1008 CTGCGTCCTTTGGAGGGAAAACTCCTCCCGCGTGCCTGGGGTAGGGCTCCACTCTTCCACGCTCC 1067

Qy      720 CAAGGACGCCCAGCATCCCCACTCGCGACAAATACGACGCGCACCGTGGGCTCAAGTTCCGG 779
            |||||||||||||||||||||||||||||||||||||||
Db     1068 CAGGAACACGCAGCATCCCCACTACGACAATACGACGTCAACGTCACGGTGGGGC 1127

Qy      780 GGCTGCCCTTCTCGTCCCGCTATGTACGTGGGGGGGATCTCTGCCGGATCGTTTTTCC 832
            ||||||||||||||||||||||||||||
Db     1128 GGCTGCTTTCTGCTCCGGCTATGTACGGTGGGGGGATCTCTGTCGGATCGTTTTTCC 1180
```

RESULT 7

AAQ040429

ID   AAQ040429 standard; cDNA; 1734 BP.
XX
AC   AAQ040429;
XX
DT   30-JUL-1993 (first entry)
XX
FH   Key   Location/Qualifiers
FT   CDS   325..1734
FT   /*tag= a
FT   /note= "not full-length CDS, hence no termination codon"
XX
PN   JP05068562-A.
XX
PD   23-MAR-1993.
XX
PF   30-MAY-1991;   91JP-00153736.
XX
PR   30-MAY-1991;   91JP-00153736.
XX
PA   (SANW) SANWA KAGAKU KENKYUSHO CO.
XX
CC   AAQ040425-Q40439
CC   sequences of these clones can be used in diagnostic assays for HCV. See
CC   amplification, including clone JK2-B. Primer/probes derived from the
CC   nucleotides long) and 14 shorter clones were isolated by PCR
CC   cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B (9405
PS   Claim 4; Page 22-24; 44pp; Japanese.
PT   DNA and cDNA of hepatitis C virus - useful as probes for diagnosing HCV
PT   infection.
XX
DR   WPI; 1993-130638/16.
DR   P-PSDB; AAR34471.
XX
DE   Hepatitis C virus clone JK2-B.
KW   HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
KW   polymerase chain reaction; diagnostic method; ds.
XX
OS   Hepatitis C virus.
XX
SQ   Sequence 1734 BP; 324 A; 540 C; 506 G; 364 T; 0 U; 0 Other;

Query Match   88.6%; Score 738.6; DB 2; Length 1734;
Best Local Similarity   94.5%; Pred. No. 3.9e-190;
Matches 787; Conservative 0; Mismatches 44; Indels 2; Gaps 2;

```
Qy     1 AGAAAAAACAAAACGTAACACCAACCTCCCCCCACAGGACGTCAGGTTCCCGGGGGGTGGT 60
         ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   349 AGAAAAAACAAAACGTAACACCAACCTCCCCCCACAGGACGTCAAGTTCCCGGGGGGTGGC 408

Qy    61 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGCGCGCCACAGGACGTCAAGTTCCCGGGGGGTGGC 120
         |||||||||||||||||||||||||||||||||||||
Db   409 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGCGCGCCACAGGACGCCCCAGGTTGGGTGTGCGCGCG 468

Qy   121 ACTAGGAAGACTTCCGAGCGGTCGGCAACCTCGTTGGAAGGCGGACAACCTATCCCCAAGGCT 180
         ||||||||||||||||||||||||||||||||||
Db   469 ACTAGGAAGACTTCCGAGCGGTCGGCAAACCTTCGTTGGGAAGGCGGACAACCTATCCCCAAGGCT 528

Qy   181 CGCCAGCCCGAGGGCGCCAGGGCCCTGGGCTTCAGCCCCGGGGTACCCTTGGCCCTCTATGGCAAT 240
         |||||||||||||||||||||||||
Db   529 CGCCAGCCCGAGGGCGCCAGGGCCCTGGGCTCAGCCCCGGGGTACCCTTGGCCCTCTATGGCAAT 588

Qy   241 GAGGGCCATGGGGGTGGGGCCAGGAGGGGCTACCCTTGGGGGTCCCCGGGGCCTAGTTGGGGC 300
         |||||||||||||||||||||||||||
Db   589 GAGGGTCTGGGGTGGGGCCAGGAGGGGCTCCCCTGTCCCCCGGGGTACCCCTTGGCCCTAGTTGGGGC 648

Qy   301 CCGACTGACCCCGCGTAGGTCGGGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360
         ||||||||||||||||||||||||||||
Db   649 CCGACGGACCCCGGGTTGGGGTAGGTCGGGTAATTTGGGGTGGGTAAGGTCATCGATACCCTCACATGC 708

Qy   361 GGCTTCGGCCGGACTCTCATGGGGGTACAATTCCGGCTCGTCGTCCCGGCTCCCCTTTAGG-GGGGCTGC 419
         |||||||||||||||||||||||
Db   709 GGCTTCGGCCGGAC-CTCATGGGGGTACAATTCCGGCTCCTCGTCCCGGCTCCCCTTTAGG-GGGGCTGC 767

Qy   420 CAGGGGCCCTGGGCCATGGGGTTCGGGGTTCTGGAGGACGGGGTGAACTATGCCAACAGGGAA 479
         |||||||||||||||||||||||||||||||
Db   768 CAGGGGCCCTGGGCCATGGGGTTCTGGAGGACGCGGGGTGAACTACGCCAACAGGGAA 827

Qy   480 TTTACCCGGTTGCTCTTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCC 539
         ||||||||||||||||||||||||||
Db   828 TCTGCCCGGTTGCTCTTTTCTCTATCTTCCTCTTGGCTCTGCGTCGTGTTTGACCATCCC 887

Qy   540 AGCTTCCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATATGCACAACGAATTGCTC 599
         |||||||||||||||||||||||||||||
Db   888 AGCATCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGACTGCTC 947

Qy   600 CAACTCAAGCCATCGTGTACAGAGACGAGGACAGATGATCATGGCCCGGGGTGTCGTGCC 659
         ||||||||||||||||||||||||||
Db   948 CAACTCAAGTATTGTGTATGAGGGCAGGCGACTTGATCATGGCCACACCCCCTGGGGTCGGTGCC 1007

Qy   660 CTGTGTCCGGGAGGGTAATTCCTCCCGCCTGTCGGGTAGCGCCCTCACTCCAAGGCT 719
         ||||||||||||||||||||||||
Db  1008 CTGGCTCCTGGAGGAAACTCCTCCCGCTGCGGGTAGCGCCTCACTCTCACGCTCGGCGGC 1067

Qy   720 CAAGGGACGCCCAGCATCCCCACCCAGGGACAATACGAGTCAGTCATGTCAGTTGCTCGTTGGTGGGGC 779
         |||||||||||||||||||||||||||
Db  1068 CAGGAACACACAGCATCCCCACTACGACGAGTCAGTCATGTCAGTTGCTCGTTGGGGC 1127

Qy   780 GGCTGCCTTCTCGTCCGGCTATGTACGGTGGGGGGATCTCTGGGGACCTGTGTTTTTCC 832
         |||||||||||||||||||||||||||
Db  1128 GGCTGCTTTCTCGTCCGGCTATGTACGGTGGGGGGATCTCTGGGGACCTGTTTTTCC 1180
```

FT   CDS   325..1734
FT   /*tag= a
FT   /note= "not full-length CDS, hence no termination codon"
XX
PN   JP05068562-A.
XX
PD   23-MAR-1993.
XX
PF   30-MAY-1991;   91JP-00153736.
XX
PR   30-MAY-1991;   91JP-00153736.
XX
PA   (SANW) SANWA KAGAKU KENKYUSHO CO.
XX
DR   WPI; 1993-130638/16.
DR   P-PSDB; AAR34472.
XX
PT   DNA and cDNA of hepatitis C virus - useful as probes for diagnosing HCV
PT   infection.
XX
PS   Claim 4; Page 24-26; 44pp; Japanese.
XX
CC   cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B (9405
CC   nucleotides long) and 14 shorter clones were isolated by PCR
CC   amplification, including clone JK2-C. Primer/probes derived from the
CC   sequences of these clones can be used in diagnostic assays for HCV. See
CC   AAQ40425-Q40439
XX
SQ   Sequence 1734 BP; 322 A; 538 C; 366 T; 0 U; 0 Other;

Query Match   88.4%;   Score 737; DB 2; Length 1734;
Best Local Similarity   94.4%;   Pred. No. 1.1e-189;
Matches   786; Conservative   0; Mismatches   45; Indels   2; Gaps   2;

RESULT 9

AAQ40438

ID   AAQ40438 standard; cDNA; 1734 BP.
XX
AC   AAQ40438;
XX
DT   30-JUL-1993   (first entry)
XX
DE   Hepatitis C virus clone JK5-B.
XX
KW   HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
KW   polymerase chain reaction; diagnostic method; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   325..1734
FT   /*tag= a
FT   /note= "not full-length CDS, hence no termination codon"
XX
PN   JP05068562-A.
XX
PD   23-MAR-1993.
XX
PF   30-MAY-1991;   91JP-00153736.
PR   30-MAY-1991;   91JP-00153736.
XX
PA   (SANW) SANWA KAGAKU KENKYUSHO CO.
XX
PR   WPI; 1993-130638/16.
DR   P-PSDB; AAR34480.
XX
PT   DNA and cDNA of hepatitis C virus - useful as probes for diagnosing HCV
PT   infection.
XX
PS   Claim 4; Page 40-42; 44pp; Japanese.
XX
CC   cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B (9405
CC   nucleotides long) and 14 shorter clones were isolated by PCR
CC   amplification, including clone JK5-B. Primer/probes derived from the
CC   sequences of these clones can be used in diagnostic assays for HCV. See
CC   AAQ40425-Q40439.
XX
SQ   Sequence 1734 BP; 546 C; 504 G; 363 T; 0 U; 0 Other;

Query Match   88.4%; Score 737; DB 2; Length 1734;
Best Local Similarity   94.4%;
Matches 786; Conservative   0; Mismatches 45; Indels   2; Gaps   2;

Qy   1 AGAAAAACCAAACGTAACACCAACCTCCGCCCCACGAGAGACTCAGGTTCCGGGCGGTGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   349 AGAAAAACCAAACGTAACACCAACCGCGCCCCACAGGACGTCAAGTTCCCGGGCGGTGGGT 408

Qy   61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCGGCCCCCAGGTTGGGTGTGCGGCGCG 120

RESULT 10

AAT12974

ID   AAT12974 standard; DNA; 2433 BP.
XX
AC   AAT12974;
XX
DT   25-SEP-1996   (first entry)
XX
DE   HCV E1 construct HCC166.
XX
KW   HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
KW   serotype; reversed phase hybridisation assay; genotype; antigen; sera;
KW   ss.
XX
OS   Hepatitis C virus.
XX
PN   WO9604385-A2.
XX
PD   15-FEB-1996.
XX
PF   31-JUL-1995;   95WO-EP003031.

XX
PR 29-JUL-1994; 94EP-00870132.
PA (INNO-) INNOGENETICS NV.
XX
PI Maertens G, Bosman F, De Martynoff G, Buyse M;
XX
DR WPI; 1996-129401/13.
XX
PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope proteins
PT - in presence of di:sulphide bond cleavage agent, to produce proteins
PT suitable for direct use in vaccines or diagnostic assays of HCV.
PS Claim 23; Fig 21; 146pp; English.
XX
CC AAT12704-T12709 and AAT12961-T12974 represent hepatitis C virus (HCV) E1
CC and E2 protein coding sequence constructs. These sequences are included
CC in vectors for the production of recombinant E1, E2, and E1/E2 proteins.
CC The recombinant proteins can then be isolated using a method of the
CC invention. In the method, the envelope proteins are purified by carrying
CC out a disulphide bond cleavage, or a reduction step with a disulphide
CC bond cleavage agent, after lysis of recombinant host cells. The
CC constructs containing the purified HCV envelope proteins can be used for
CC vaccinating humans against HCV, for in vitro detection of HCV antibodies
CC in a sample, and in a serotyping assay for detecting one or more
CC serological types of HCV present in a biological sample. The constructs
CC can also be immobilised on a solid substrate and incorporated into a
CC reversed phage hybridisation assay for determining the presence or the
CC genotype of HCV. The new purification method preserves the conformation
CC of the recombinantly expressed E1, E2 and E1/E2, and eliminates
CC contaminating proteins. Antigens isolated using this method are more
CC reactive with human sera than those isolated by known techniques
XX
SQ Sequence 2433 BP; 434 A; 745 C; 714 G; 540 T; 0 U; 0 Other;

Query Match 88.4%; Score 737; DB 2; Length 2433;
Best Local Similarity 94.4%; Pred. No. 1.2e-189;
Matches 786; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

```
Qy       1 AGAAAAACAAAACGTAACACCAACCTCCGCCCACAGGAGACGTCAGGTTCCCGGGCGTGGT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 AGAAAAACAAAACGTAACACCAACCTCCGCCCACAGGAGACGTCAAGTTCCCGGGCGTGGT 84

Qy      61 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGGCCAGGGGCCCCAGGTTGGGTGTGGCGCCG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGGCCAGGGGCCCCCAGCGTTGGGTGTGGCGCG 144

Qy     121 ACTAGGAAGAACTTCCGAGCGCGTCGCAACCTCGTGAAGAGGCGACAACCTATCCCAAGGCT 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 ACTAGGAAGAACTTCCGAGCGGTCGCAACCTCGTGGGAGGCGGCACAACCTATCCCAAGGCT 204

Qy     181 CGCCAGCCCCGGAGGGCCAGGGCCCTGGGCCTCAGCCCCCGGGTACCCCTTGCCCCTCTATGGCAAC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 CGCCGACCCCAGGGTPAGGGCCTGGGCCTCAGCCCCCGGGTACCCCTCTATGGCAAT 264

Qy     241 GAGGGCCATGGGGGTGGGGCCAGGATGGGCCTCCGGTGGGCTCCCGGCCCCTAGTTGGGGC 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 GAGGGCCATGGGGGTGGGGCCAGGGATGGCCTCTCGGCCCCTAGTTGGGGC 324

Qy     301 CCCACTGACCCCCGCGTAGGTCGGTCGGTGAATTGGGTAAAAGTCATGATACCCTCACATGC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 CCTACAGACCCCCGGCGTAGGTCGGGTAATTGGGTCATCGATACCCTTACATGC 384

Qy     361 GGCTTCGCGCGACTCTCATGGGGGTACATTCCGCTCGTCGTCGCGCTCCCTTA-GGGCGCGCTGC 419
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     385 GGCTTCGCGCCAC-CTCGTGGGGTACATTCCGCTCGTCGGCCCCCTAGGGCGGGCCTGC 443

Qy     420 CAGGGCCCCTGGCGCCATGGCCGTTCTGGAGGACGGGGCTGAACTATGCAACAGGGAA 479
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     444 CAGGGCCCCTGGGCGCATGGCGGTCTCCGGGGTTCGGAGGACGGGCCGTGAACTATGCAACAGGGAA 503

Qy     480 TTTACCCGGGTTGCTCTTTCTCTATCTTCCTTGGCTTTTGTGTTCCTGTTTTGTGTTGACCGTTCC 539
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     504 TTTGCCCGGGTTGCTCTTTCTCTATCTTCCTCTGGCTTTGCTTGTCGTTGTCGTTGTTGACCGTTCC 563
```

RESULT 11
ADD55557
ID ADD55557 standard; DNA; 2433 BP.
XX
AC ADD55557;
XX
DT 15-JAN-2004 (first entry)
XX
DE Hepatitis C virus E1/E2 protein coding sequence #14.
XX
KW Hepatitis C virus; HCV; vaccine; liver disease; E1 protein; E2 protein;
KW liver fibrosis; ds; gene.
XX
OS Hepatitis C virus.
XX
PN WO200305191202-A2.
XX
PD 26-JUN-2003.
XX
PF 18-DEC-2002; 2002WO-EP014480.
XX
PR 18-DEC-2001; 2001US-00020510.
PR 16-OCT-2002; 2002US-04183358P.
XX
PI Maertens G, Depla E, Bosman F;
XX
PA (INNO-) INNOGENETICS NV.
XX
DR WPI; 2003-541632/51.
DR P-PSDB; ADD55558.
XX
PT New hepatitis C virus (HCV) vaccine composition, useful for reducing
PT liver disease, e.g., liver fibrosis in a chronic HCV-infected mammal.
XX
PS Example 2; SEQ ID NO 49; 271pp; English.
XX
CC The invention comprises an Hepatitis C virus (HCV) vaccine for reducing
CC liver disease. The vaccine of the invention comprises an HCV E1 or E2
CC protein as an antigen. The HCV vaccine is useful for reducing liver
CC disease (e.g. liver fibrosis) in a chronic HCV-infected mammal. The
CC present DNA sequence encodes an HCV E1/E2 protein.
XX
SQ Sequence 2433 BP; 434 A; 745 C; 714 G; 540 T; 0 U; 0 Other;

Query Match 88.4%; Score 737; DB 10; Length 2433;
Best Local Similarity 94.4%; Pred. No. 1.2e-189;
Matches 786; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

```
Qy       1 AGAAAAACAAAACGTAACACCAACCTCCGCCCACAGGAGACGTCAGGTTCCCGGGCGTGGT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   504 TTTGCCCGGGTTGCTCTTTCTCTATCTTCCTCTGGCTTTGCTTGTCGTTGTCGTTGTTGACCGTTCC 563
Qy   540 AGCTTCCGCTTATGAAGTGCGGCAAACGGTGTCCGGGGATCTAACCATTGTCACGGAAGGATTGCTC 599
Db   564 AGCTTCCGCTTATGAAGTGCGGCAAACGGTGTCCGGGGATCTAACCATTGTACCATGTAACGGAACGGACTTGCTC 623
Qy   600 CAACTCAAGCATCTGTACACGAGACAGCGGACAGATGATCATCACAGCCCCCGCGGTGTGTGGCC 659
Db   624 CAACTCAAGCATTGTGTTATGAGGGCAGCGGACAGATGATCATCGACAACACCCCCGCGGTGCGTGCC 683
Qy   660 CTGTGTCCGGGAGGGTAATTCCTCCCGGCTGCTTGGGGTCGTGAGGGTGTCACTCCTGCCTCCCGGC 719
Db   684 CTGCGTTCGGGAGGAGAACAAACTCTTCCCGCTGCTGGGGGGCGCTCACCCCAGGGTGCGGGGC 743
Qy   720 CAAGGACGGCCAGCATCCCACTGGCAGCAATACGACGTGCACGACGTGCGATTTGCTGCTGTGAGG 779
Db   744 TAGGAACGGCCAGGCGTCCCACCACGACAATACGACCGCACCTCGATTTGCTGCTGTGGGGC 803
Qy   780 GGCTGCGTTCCGTTCGCGTTCCGGCTATGTAGGTACGTGGGGGGATCTCTTCGCCGGGGATCTGTTTTCC 832
Db   804 GGCTGCGTTTCTGTTTCCGGCTATGTACGTGGGGGGGGGGGACCTTCGCGGATCGTCTTCTTCC 856

```
Db    25  AGAAAAACCAAACGTAACCAACCGCCCCCCACCAGGAGGAGTCAAGTTCCCGGGGCGGGGT   84

Qy    61  CAGATCGTTGGTGGAGTTTACCTTGTGCCGCGCGCAGGGGGCCCCAGGTTGGGTGTGCGCCG  120
              ||||||||||||||||| ||||| |||||||||| ||||||||  || |||| ||| |||
Db    85  CAGATCGTTGGTGGAGTTTACCGTTGCCGCGCGCAGGGGGCCCCAGGTTGGGTGTCGCGCG  144

Qy   121  ACTAGGAAGAGACTTCCGAGCGGGTCGCCAACCTCGTGTGGAGGCGACAAACCTATCCCCAAGGCT  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145  ACTAGGAAGAGACTTCCGAGCGGGTCGCCAACCTCGTGTGGAGGCGACAAACCTATCCCCAAGGCT  204

Qy   181  CGCCAGCCCGAGGGCAGGCAGGGCCTGGGCTCAGCCCCGGGTACCCTTCTATGGGCAAC  240
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   205  CGCCGACCCGAGGGTAGGGCCCTGGGCTCAGCCCCGGGTACCCTTCTATGGGCAAT  264

Qy   241  GAGGGGCCATGGGGTCGGCAGGAGTGGCTCTCTTGCTCACCCCGGGTCGGCC...  300
              ||||||||||||||||||||||||||||||||||||||||||||||
Db   265  GAGGGCCATGGGGTCGGCAGGAGTGGCTCCTCTTGCTCACCCCGGGTCGGCC...  324

Qy   301  CCCACTGACCCCCCCGTAGGTCCGTAATTGGGTAAAGTCAATGATACCCTAGTTGGGGC  360
              |||||| ||||||||||  ||||||||||||||||||||| |||||||||||||||||
Db   325  CCTACAGACCCCCGGGTAGGTGCGGTAATTGGGTAAGGTCATCGATACCCTTCACATGC  384

Qy   361  GGCTTCCGCGACTCTCATGGGGTACATTCCGCTCGTCGTCCCTTA-GGGGGCGCTGC  419
              ||| ||||||| |||||||||||||||||||||||||||||||| ||||||| ||||
Db   385  GGCTTCGGCGAC-CTCGTGGGGGTACATTCCGGCTCGTCGGGGCGCGTGC  443

Qy   420  CAGGGCCCCTGGGCCCATGGCGTCTCGGGGTTCTGGAGGAGCGCGGGTGAACTATGCAACAGAGGAA  479
              |||||||||||||||| |||||||| ||||||||||||||||||||||| ||||||||||||||||
Db   444  CAGGGCCCTGGGCCGGCATGGCGTCCCGGGGTTCTGGAGGACGGCGTCGAACTATGCAACACAGGAA  503

Qy   480  TTTACCCCGGTTGCTCTTTCTCTATCTTCCCTGCTTGGCTTTGCTGTTTGACCAATTCC  539
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   504  TTTGCCCCGGTTGCTCTTCTTCTCTTCCCGCTTGGCTTTGCTGTCTGTTTGACCAGGTTCC  563

Qy   540  AGCTTCCCGCTTATGAAGTGCGCAAACGTGTCCCGGGATCTACCACCATGTCACGGAACCATTGCTC  599
              ||||||||||  ||||||||||||||||||| ||||||||||||||||| |||||||||| |||||
Db   564  AGCTTCCCGCTTATGAAAGTGCGCAACGTCCGGGGATGTACCATGTGGTCACGGAACACTGCTC  623

Qy   600  CAACTCAAGCACATCGTGTACGAGACGACGGCGACATGATCATGCACCACCCCGGGTGTGTGGTGCC  659
              ||| |||||||||| ||||| || |||||||||||||||| |||||||||||||||||||||||
Db   624  CAACTCAAGGCCTTATGAGAGTCGCGACAGTGTCCGGGATCTTACCACCGTGCCGCACAGCTGCTC  683

Qy   660  CTGTGTCCGGGAGGGTAATTCCTCCCGGCGTGCTGGGGGTAGGCGCTCACTCCCACGGTCGCGGC  719
              ||||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   684  CTGCGGTTCGGGAGGAAACAACTCTTCCCGGCGTGCTGGGGTAGGCGCTCACTCCCACGGTCGCGGC  743

Qy   720  CAAGGACGGCCAGCATCCCCACTCGCACAATACGACGCCACGGTCCATTTGCTCGTTGTGGGGGC  779
              | |||||||||| |||||||||||||||||||||||||||| ||||||||||||||||||
Db   744  TAGGAAGCGCCAGGCGTCCCCCACGACAATACGACGCCACGTGCGATTTGCTCGTTGTGGGGC  803

Qy   780  GGCTGCCTTTCTCGTCCCGCTATGTTACGTCGGGGGGATCTCTGCGGATTCTGTTTTTCC  832
              | |||| |||||||||||||||||||||||||| ||||||||||||||
Db   804  GGCTGCTTTTCTCGTTCCGCTATGTACGTGGGGGACTTCTGCGGANCTGTGCTTCC  856
```

CC   The invention relates to the use of a hepatitis C virus (HCV) vaccine
CC   composition for reducing liver disease (such as liver fibrosis or its
CC   progression), serum alanine aminotransferase (ALT) levels, steatosis, or
CC   anti-E2 immunoreactivity in the liver of a chronic HCV-infected mammal,
CC   or for treating a chronic HCV-infected mammal. The liver disease is
CC   reduced by at least 1-2 points according to the overall Ishak score in
CC   the HCV-infected mammal. Also included are a method for predicting
CC   changes in liver disease in a chronic HCV-infected mammal, a therapeutic
CC   HCV vaccine composition (comprising at least one purified or a
CC   combination of at least 2 HCV single or specific oligomeric recombinant
CC   envelope protein selected from an E1 or E2 protein, a part of E1 and E2
CC   protein, an E1/E2 protein complex formed from purified HCV single or
CC   specific oligomeric recombinant E1 or E2 proteins or its parts and
CC   optionally a pharmaceutical adjuvant), a composition (comprising at least
CC   one E1 or E2 peptide, and optionally, a pharmaceutical adjuvant), an
CC   immunogenic HCV composition (or HCV vaccine composition), comprising a
CC   recombinant virus allowing expression of at least one HCV recombinant
CC   envelope protein (selected from an E1 protein and/or an E2 protein, and
CC   their parts, and optionally, a pharmaceutical adjuvant) and an HCV
CC   vaccine composition (comprising a recombinant envelope protein chosen
CC   from an E1 protein and/or an E2 protein, and parts of the E1 and E2
CC   protein and/or an E2 protein, and parts of the E1 and E2 proteins and,
CC   optionally, a pharmaceutical adjuvant. The HCV vaccine composition is
CC   useful for reducing liver disease (such as liver fibrosis or its
CC   progression), serum ALT levels, steatosis, or anti-E2 immunoreactivity in
CC   the liver in a chronic HCV-infected mammal, or for treating a chronic HCV
CC   -infected mammal, particularly human. The HCV E1 proteins are useful for
CC   in vitro monitoring HCV disease or prognosing the response to treatment
CC   of patients suffering from HCV infection. The present sequence is an HCV
CC   cDNA encoding an E1 protein (or fragment).

XX

SQ   Sequence 2433 BP; 434 A; 745 C; 714 G; 540 T; 0 U; 0 Other;

Query Match   88.4%; Score 737; DB 12; Length 2433;
Best Local Similarity   94.4%;
Matches   786; Conservative   0; Mismatches   45; Indels   2; Gaps   2;

```
Qy     1  AGAAAAACCAAACGTAACACCAACCTCCCGCCCCCCACCAGGAGGAGTCAGGTTCCCGGGGCGGGGT   60
              |||||||||||||||| ||||||  |||||||||||||||||||||||||  |||||||||||||
Db    25  AGAAAAACCAAACGTAACCAACCGCCCCCCACCAGGAGGAGTCAAGTTCCCGGGGCGGGGT   84

Qy    61  CAGATCGTTGGTGGAGTTTACCTTGTGCCGCGCGCAGGGGGCCCCAGGTTGGGTGTGCGCCG  120
              ||||||||||||||||| ||||| |||||||||| ||||||||||| |||| ||| |||
Db    85  CAGATCGTTGGTGGAGTTTACCGTTGCCGCGCGCAGGGGGCCCCAGGTTGGGTGTCGCGCG  144

Qy   121  ACTAGGAAGAGACTTCCGAGCGGGTCGCCAACCTCGTGTGGAGGCGACAAACCTATCCCCAAGGCT  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145  ACTAGGAAGAGACTTCCGAGCGGGTCGCCAACCTCGTGTGGAGGCGACAAACCTATCCCCAAGGCT  204

Qy   181  CGCCAGCCCGAGGGCAGGCAGGGCCTGGGCTCAGCCCCGGGTACCCTTCTATGGGCAAC  240
              |||| ||||||||| || ||||||||||
Db   205  CGCCGACCCGAGGGTAGGGCC...
```

```
RESULT 13

ID   ADR82189 standard; DNA; 9587 BP.
XX
AC   ADR82189;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Hepatitis C virus type 1b polyprotein DNA.
XX
KW   antilipemic; cardiant; vasotropic; antiarteriosclerotic; antidiabetic;
KW   cytostatic; anticonvulsant; nootropic; muscula; anti-HIV;
KW   RNA interference; iRNA; antisense technology; lipid metabolism;
KW   cholesterol imbalance; dyslipidaemia hypercholesterolaemia;
KW   coronary artery disease; CAD; coronary heart disease; CHD;
KW   atherosclerosis; hepatic glucose production;
KW   glucose-metabolism-related disorder; diabetes; cancer; breast cancer;
KW   colon cancer; lung cancer; neurological disease; Huntington disease;
KW   spinocerebellar ataxia; viral disease; AIDS; HCV; polyprotein; gene; ds.
OS   Hepatitis C virus.
XX
PN   WO2004080406-A2.
XX
PD   23-SEP-2004.
XX
PF   08-MAR-2004; 2004WO-US007070.
XX
PR   07-MAR-2003; 2003US-0452682P.
PR   12-MAR-2003; 2003US-0454265P.
PR   13-MAR-2003; 2003US-0454962P.
PR   13-MAR-2003; 2003US-0455050P.
PR   14-APR-2003; 2003US-0462894P.
PR   17-APR-2003; 2003US-0463772P.
PR   25-APR-2003; 2003US-0465665P.
PR   25-APR-2003; 2003US-0465802P.
PR   09-MAY-2003; 2003US-0469612P.
PR   08-AUG-2003; 2003US-0493986P.
PR   11-AUG-2003; 2003US-0494597P.
PR   26-SEP-2003; 2003US-0506341P.
PR   09-OCT-2003; 2003US-0510246P.
PR   10-OCT-2003; 2003US-0510318P.
PR   07-NOV-2003; 2003US-0518453P.
XX
PA   (ALNY-) ALNYLAM PHARM.
XX
PI   Manoharan M, Bumcrot D;
XX
DR   WPI; 2004-677362/66.
XX
PS   Example 5; SEQ ID NO 6688; 378pp; English.
PT   Interference RNA agent useful for treating dyslipidemias, coronary artery
PT   disease, diabetes, cancer or neurological disease, comprises sense
PT   sequence and antisense sequence which has specific modifications.
XX
CC   The invention describes a RNA interference (iRNA) agent (I) comprising a
CC   sense sequence and an antisense sequence, where the sense sequences have
CC   one or more asymmetrical 2'-O alkyl modifications, the antisense
CC   sequences have one or more asymmetrical phosphorothioate modifications
CC   and the antisense sequence targets a human gene sequence. Also described
CC   are: a pharmaceutical preparation comprising (I); reducing (M1) apoB-100
CC   levels or glucose-6-phosphatase levels in a subject; producing (I);
CC   stabilising (I), involves selecting a sequence with activity and
CC   introducing one or more asymmetrical modification in the sequence, where
CC   the modification decreases nuclease sensitivity while not decreasing its
CC   activity; a kit comprising (I) and instruction for its use; and a device
CC   that can be dispense or administer a composition comprising (I). (I) is
CC   useful for reducing apoB-100 levels or glucose-6-phosphatase levels. (M1)
CC   is useful for reducing apoB-100 levels or glucose-6-phosphatase levels.
CC   The subject is suffering from a disorder characterised by elevated or
CC   otherwise unwanted expression of apoB-100, elevated or otherwise unwanted
CC   levels of cholesterol, and/or dysregulation of lipid metabolism. The
CC   disorder is chosen from the HDL/LDL cholesterol imbalance,
CC   dyslipidaemias, hypercholesterolaemia, statin-resistant
CC   hypercholesterolaemia, coronary artery disease (CAD), coronary heart
CC   disease (CHD) and atherosclerosis. (I) is administered to a subject to
CC   inhibit hepatic glucose production or for treating glucose-metabolism-
CC   related disorder e.g. diabetes or type-2 diabetes. (I) is useful for
CC   treating the diseases as mentioned above, cancer (e.g. breast, colon or
CC   lung cancer), neurological disease (e.g. Huntington disease or
CC   spinocerebellar ataxia) or viral disease (e.g. AIDS). This sequence
CC   represents hepatitis C virus type 1b polyprotein DNA.
XX
SQ   Sequence 9587 BP; 1921 A; 2865 C; 2714 G; 2087 T; 0 U; 0 Other;

Query Match             88.4%;  Score 737;  DB 13;  Length 9587;
Best Local Similarity   94.4%;
Matches  786;  Conservative    0;  Mismatches   45;  Indels    2;  Gaps     2;
                                    Pred. No. 1.8e-189;

Qy         1 AGAAAAACCAAAACCGTAAACACCAAACCTTCCCCCACCCACCAGGGACGTTCAGGTTCCCGGGGCGTTGGT 60
Db       366 AGAAAAACCAAACCGTAACACCAACCGCCCACCACCAGGGACGTTAAGTTCCCGGGGCCGTTGGT 425

Qy        61 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGCGCAGGGGCCCCAGAGTTGGGTTGGGCGCGCGCG 120
Db       426 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGCGCCACAGGACGTTAAGTTCCGGGGGTTGCGCGCG 485

Qy       121 ACTAGGAAAGACTTCCGAGCGCGTCGCAACCTCGTGGAAGGCGACAACCTATCCCAAGGCT 180
Db       486 ACTAGGAAAGACTTCCGAGCGCGTCGCAACCTCGTGGAAGGCGACAACCTATCCCAAGGCT 545

Qy       181 CGCCAGCCCCGAGGGCGCAGGGCGCCTGGGCTCAGCCCCGGGTACCCTTGGCCCTCTATGGCAAC 240

Db       205 CGCCGACCCGAGGGGTAGGGCCTGGGCTCAGCCCCGGGTACCCTTGGCCCTCTATGGCAAT 264

Qy       241 GAGGGGCATGGGGTGGGGCAGGATGGCTCCTTGTCACCCGGGTCTCCCGGGGCCTAGTTGGGC 300
Db       265 GAGGGGCATGGGGTGGGGCAGGATGGCTCCTTGTCACCCGGGTCTCCCGGGGCCTAGTTGGGC 324

Qy       301 CCCACTGACCCCGGGCGTAGGTCGCGTAATTGGGGTAAAGTCATCGATACCCTCACATGC 360
Db       325 CCTACAGACCCCCGGGCGTAGGTCGCGTAATTTGGGTAAGGTCATCGATACCCTTACATGC 384

Qy       361 GGCTTCGCCGGACTCTCATGGGGTACATTCCGCTCGTCTTCGATACCCTTA-GGGCGCGGTGC 419
Db       385 GGCTTCGCCGAC-CTCGTGGGGTACATTCCGCTCGTCTCGGTACCCTTACATGC 443

Qy       420 CAGGGCCCCTGGGGCCATGGGGCTCCGGGGTTCTGGAGGACGGGGGGAACATATGCAACAGGGGAA 479
Db       444 CAGGGCCCTGGGCGCGATGGGGTCCGGGTTCTGGAGGACGGGTGAACTATGCAACAGGGGAA 503

Qy       480 TTTACCCGGTTGCTCTTTCCTCTACTTCCCTCTGGCTTTTGTCTGTCGTTGTTGACCCATTCC 539
Db       504 TTTGCCCGGTTGCTCTTTCTCTACTTCCCTCTTGGCTTTGTCTTTGTTGACCCATTCC 563

Qy       540 AGCTTCCGCTTATGAAGTGCGGCGGAAGTGTTCCGGGGCAAGTGTCCAGGAGTGTCACACGAATTGCTC 599
Db       564 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGGATGTGTCACACGAATTGCTC 623

Qy       600 CAACTCAAGCATCGTGTACGGAGACACAGCGGACACTGATGATCATCACACCCCGGGTGGTGTGCC 659
Db       624 CAACTCAAGCCATCTGTATGAGGGACAGCATGATCATCACACCCCCGGGTGCCGTGCC 683

Qy       660 CTGTGTCCGGGAGGGTAATTCCTCCCGGCGTGCTGGGTAGGCGCCTCACTCGTCGCGCGGC 719
Db       684 CTGCGTTCGGGAGGAACAACTCTTCCCGCGTGGGTAGCCCTCACCCCACGCGTCGGCAGC 743

Qy       720 CAAGGACGCGGCCAGCATCCCCACTCGGGACAACAATACGGACGCCCACGTCGAATTCGCTCGTTGGGGC 779
Db       744 TAGGAAACGGCCAGGGTCCCCACCACGACGACACACGTCGAATTTGCTCGTCGTTGGGGC 803

Qy       780 GGCTGGCCTTCGTGTCCGCTAGTGTAGTGCGGGGATCTCTGGGGATCTGTTTTCC 832
Db       804 GGCTGGCTTTCTTGTTCCGCTATGTAGTGCGGGGACCTCTGCGGATCTGTGCTTCC 856
```

PA   (GLDS) LG CHEM LTD.
XX
PI   Choo SH, Lee IH, Ryoo WS;
XX
DT   WPI; 1998-492654/42.
DR   P-PSDB; AAM47264.
XX
PT   Cuticle protein 1 and 2 secreting hepatitis C virus (Japanese).
XX
PS   Disclosure; Page 2-4; 7pp; Korean.
XX
CC   The present invention relates to cuticle protein 1 and 2 secreting
CC   hepatitis C virus. The present sequence is a coding sequence provided in
CC   the exemplification of the invention
XX
SQ   Sequence 2187 BP; 406 A; 669 C; 631 G; 481 T; 0 U; 0 Other;

Query Match   88.0%;   Score 733.8; DB 2;   Length 2187;
Best Local Similarity   94.1%;   Pred. No. 8.4e-189;
Matches 784;   Conservative 0;   Mismatches 47;   Indels 2;   Gaps 2;

RESULT 14
ID   ABA03491 standard; DNA; 2187 BP.
XX
AC   ABA03491;
XX
DT   15-MAR-2002 (first entry)
XX
DE   Cuticle protein 1 and 2 secreting hepatitis C virus related DNA #1.
XX
KW   Cuticle protein 1; cuticle protein 2; hepatitis C virus; ds.
OS   Unidentified.
XX
FH   Key   Location/Qualifiers
FT   CDS   1. 2187
FT   /*tag=   a
FT   /product= "AAM47264"
FT   /partial
FT   /note= "no stop codon"
XX
PN   KR97065713-A.
XX
PD   13-OCT-1997.
XX
PF   19-MAR-1996;   96KR-00007404.
XX
PR   19-MAR-1996;   96KR-00007404.
XX

```
Qy    780 GGCTGCCTTCTCGTGTCCGCTATGTAGTGACGTGGGGGGATCTCTGTTTTTCC 832
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    804 GGCTGCCTTCTCGCTCCGCTATGTACGTGGGGGATCTCTGGGATCTGTTTTCC 856
```

RESULT 15

```
Qy     ?
Db
```

ID   AAQ43889 standard; cDNA to mRNA; 2540 BP.
XX
AC   AAQ43889;
XX
DT   21-OCT-1993 (first entry)
XX
DE   NANB hepatitis virus polynucleotide N-2540-2.
XX
KW   Non-A, non-B; virus; polymerase chain reaction; detection; sensitive;
KW   specific; HCV; NANBH; ss.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
FH   Key   Location/Qualifiers
FH
FT   5'UTR   1..341
FT   /tag= b
FT   CDS   342..2540
FT   /tag= a
FT   /note= "from 5' terminal of NANBH virus RNA"
XX
PN   JP05091884-A.
XX
PD   16-APR-1993.
XX
PF   10-APR-1991; 91JP-00196175.
XX
PR   12-JUN-1990; 90JP-00153401.
PR   08-NOV-1990; 90JP-00304405.
XX
PA   (NAKA/) NAKAMURA T.
XX
DR   WPI; 1993-199637/25.
DR   P-PSDB; AAR38279.
XX
PT   Antigen related to non-A and non-B hepatitis virus - comprises non-
PT   translation region comprising 340 - 341 mols. of nucleotides, non-
PT   translation region comprising 1885 - 2551 mols. of nucleotides including
PT   region 1,149 and, etc.
XX
PS   Claim 3; Page 19-20; 73pp; Japanese.
XX
CC   The sequence is that of NANB hepatitis virus polynucleotide N-2540-2
CC   which codes for a non-A, non-B (NANB) hepatitis virus gene HC-OM. The
CC   polypeptide it encodes may be used in a system for detecting NANB
CC   hepatitis. This method is highly specific and sensitive, and can detect
CC   NANB hepatitis virus which could not be detected by conventional methods
XX
SQ   Sequence 2540 BP; 471 A; 775 C; 741 G; 553 T; 0 U; 0 Other;
```

```
Qy      1 AGAAAAACCAAACGTAACACACCAACCTCCGCCACAGGAGGTCAGGTTCCGGACGTGGGT 60
          ||||||||||||| ||  || |||||| |||||||||||||||||||||||||||||||||
Db    366 AGAAAAACCAAACGTAACACCAACCGCCCCCCCCCCACAGGAGGTCAAGTTCCCGGACGTGGGT 425
```

```
Qy     61 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGCGCAGGGGCCCCAGGTTGGGTGTCGCGCG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 CAGATCGTTGGTGGAGTTTACCTGTTGCCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 485
```

```
Qy    121 ACTAGGAAGAGACTTCCGAGCGGGGGGTCGCAACCTCGTTGGAAGGCGACAACCTATCCCCAAGGCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    486 ACTAGGAAGAGACTTCCGAGCGGGGGTCGCAACCTCGTTGGAATGGCGACAACCTATCCCCAAGGCT 545
```

```
Qy    181 CGGCAAGCCCCGAGGGGCGAGGGGCCTGGGCCTCACCCCGTACCCCGTCGCTTGCCCCTTC 240
          ||  ||||||||  |||||||||||||||| |||||||||||||||||||||||||||||
Db    546 CGCCCGGACACCCGAGGGGCCAGGGGCCTGGGCCTCAGCCCGTACCCCGTACCCTCTAGTTGGGC 605
```

```
Qy    241 GAGGGGCCATGGGGGGTGGGGCAGGATGGCTTCCTGTCTCGTCACCCCGTGGCCTCCCGTGCCCCGTCTCCGGTTA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 GAGGGGCCTTGGGGTGTGGGGCCAGGATGGCTTCCTGTCTCACCCCGTACCCCGTGGGGC 665
```

```
Qy    301 CCCACTGACACCCCCGCGGGGTAGGTCGGGTAACTTGGGGTAAAGTCAGATCATCGAGATCACCATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 CCCACGGGACCCCCGCGGGGTAGGTCGGGTAACTTGGGGTAAGGTCATGATACCCTTACATGC 725
```

```
Qy    361 GGCTTCGGCCGACTCTCATGGGGTACATTCCGGTCTGTGGGGCTCCCCGTGCGGGGGTGCGTGCTGC 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    726 GGCTTCGGCCGA-TCTCATGGGGTATATTCCCGTCTTCGGGGGTCATGATACCCTTAGGTCATGC 784
```

```
Qy    420 CAGGGGCCTCGGCGTCCGGGTTCTGGGAGGACGGGGTGAACTATCCAACAGGGAA 479
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    785 CAGGGGCCTCGGCTCACGGGTGTCCGGGGTTCTGGGAGGACGGGGTGAACTATCCAACAGGGAA 844
```

```
Qy    480 TTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCGTTTGCTGTGTCCTGTTTGACCATTCC 539
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    845 CTTTGCCCGGTTGCTTCTTTTCTCTTATCTTCCTCTTGGCGTTTGCTGTCGTCCGTGTTTGACCATCCC 904
```

```
Qy    540 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGGATCTACCATGTCACGAACGATTGCTC 599
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    905 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGGATATACCATGTCACGAACGACTGCTC 964
```

```
Qy    600 CAACTCAAGCATCGTGTTACGAGACAGCAGGACAGATCATGCAACACCCCCCGGGGTGTGCC 659
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    965 CAACTCAAGCATTGTGTTATGAGGCAGCGGACAATGCATCAACTACTCCCCGGGGTGGGCC 1024
```

```
Qy    660 CTGTGTCCGGCGAGGGTAATTCCTCCCGCTGCTGTGGGTAGCGCTCACTCCACGCTGCGGC 719
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1025 CTGCGGTTCGGGGAGGACACAGCTCCCGTTGCTGGGGCTCACTCCCAGGCTCGCGGC 1084
```

```
Qy    720 CAAGGACGCGCAGCCATCCCCACTGGGCACAATTACGACGCCACGTCGCATTTGCTTGCTTGGGGGC 779
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1085 CAGGAATGCCACGGTCCCCACTACGACAAATACGACCCCAGGTCGACTTGCTTCCGTTGGGGC 1144
```

```
Qy    780 GGCTGCCTTCTCGTGTCCGCTATGTACGTGGGGGGATCTCTGGGGATCGTTTTCC 832
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1145 GGCTGCCTTCTCGCTCCGCTATGTACGTGGGGGATCTCTGGGGATCTGTTTTTCC 1197
```

Query Match   88.0%; Score 733.8; DB 2; Length 2540;
Best Local Similarity 94.1%;
Matches 784; Conservative 0; Mismatches 47; Indels 2; Gaps 2;
Pred. No. 8.8e-189;

Search completed: February 21, 2005, 01:56:31
Job time : 595 secs

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: February 21, 2005, 01:13:20 ; Search time 3559 Seconds
                                        (without alignments)
                                        8919.809 Million cell updates/sec

Title: US-09-664-363-5
Perfect Score: 834
Sequence: 1 AGAAAAACCAAACGTAACAC..........TCTGCGGATCTGTTTTCCCG 834

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
                1: gb_est1:*
                2: gb_est2:*
                3: gb_htc:*
                4: gb_est3:*
                5: gb_est4:*
                6: gb_est5:*
                7: gb_est6:*
                8: gb_gss1:*
                9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB	ID	Description
1	57.4	6.9	1100	9	CNS016KD	AL106855 Drosophil
2	56.2	6.7	925	9	CNS001P	AL053013 Drosophil
3	53.4	6.4	879	9	CNS02QOG	AL209545 Tetraodon
4	48.6	5.8	1101	9	CNS016UE	AL107216 Drosophil
5	48	5.8	1101	9	CNS00BNG	AL057398 Drosophil
6	47.8	5.7	1036	9	CNS010BS	AL098770 Drosophil
7	47.6	5.7	1030	9	AG126333	Pan trogl
8	45.8	5.5	925	9	CNS001P	AL053013 Drosophil
9	44.8	5.5	844	9	CNS0052P	AL056652 Drosophil
10	44.8	5.4	910	9	CNS006ON	AL055629 Drosophil
11	44.4	5.3	439	7	CO141650	CO141650 EST836321
12	44.4	5.3	450	7	CO137934	CO137934 EST832605
13	44.4	5.3	743	7	CO138520	CO138520 EST833191
14	43.6	5.2	548	7	CO143444	CO143444 EST838115
15	43.6	5.2	899	9	CD109759	CD109759 AGENCOURT
16	43.2	5.2	932	9	CNS0072Q	CD109759 AGENCOURT
17	43	5.2	1201	9	CNS016BR	AL106545 Drosophil
18	42.6	5.1	932	9	CNS0072Q	AL066742 Drosophil
19	42.6	5.1	936	4	BG852371	BG852371 1024034A0
20	42	5.0	400	7	CO137412	CO137412 EST832083
21	41.8	5.0	442	1	AF367693	AF367693
22	41.6	5.0	692	9	CNS007WH	AL050923 Drosophil
23	41.6	5.0	839	9	CNS004NB	AL054280 Drosophil
24	41.6	5.0	1135	5	BU527635	BU527635 AGENCOURT
25	41.2	4.9	387	7	CO148295	CO148295 EST823348
26	41.2	4.9	1009	9	CNS010EW	AL098882 Drosophil
27	40.8	4.9	518	9	CNS016UP	AL107227 Drosophil
28	40.8	4.9	997	9	CNS006DN	AL065132 Drosophil
29	40.6	4.9	980	9	AG071642	AG071642 Pan trogl
30	40.4	4.8	1203	9	CNS015Y4	AL106054 Drosophil
31	40.4	4.8	1233	6	CA975828	CA975828 AGENCOURT
32	40.2	4.8	405	7	CO893216	CO893216 BovGen_21
33	40.2	4.8	466	2	AW670394	AW670394 114401_MA
34	40.2	4.8	548	7	CK952578	CK952578 4091981_B
35	40.2	4.8	781	7	CO876886	CO876886 BovGen_05
36	39.6	4.8	801	7	CK847989	CK847989 970689_MA
37	40	4.8	842	9	AG058791	AG058791 Pan trogl
38	40	4.7	933	5	BX383247	BX383247 Drosophil
39	39.6	4.7	1014	6	CB204069	CB204069 AGENCOURT
40	39.4	4.7	356	5	BY235468	BY235468 Mus muscu
41	39.4	4.7	750	9	AG425278	AG425278 Drosophil
42	39.4	4.7	798	5	BP144487	BP144487 Mus muscu
43	39.4	4.7	869	9	AG162064	AG162064 Pan trogl
44	39.4	4.7	1101	9	CNS017SY	AL108460 Drosophil
45	39.2	4.7	761	9	CG373319	CG373319 OG1B14TV

ALIGNMENTS

RESULT 1
CNS016KD
LOCUS   CNS016KD   1100 bp   DNA   linear   GSS 26-JUL-1999
DEFINITION CNS016KD Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN16D22 of DroBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106855
VERSION   AL106855.1 GI:5624152
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1100)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT   - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES   Location/Qualifiers
     source   1..1100
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN16D22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

Query Match   6.9%; Score 57.4; DB 9; Length 1100;
Best Local Similarity   21.9%;
Matches 114; Conservative 171; Mismatches 236; Indels 0; Gaps 0;

Qy   25 CTCCCGCCCACAGGACGTCAGGTTCCCGGGCGGGGGCGGTGGTCAGATCGTTGGTGAGTTTACCTG 84
Db   556 CGCTGYCMGCGTGGKGKGKGKGKTGGGGCKGKSGTSGGGGSGCGCTGG 615

```
Qy    85 TTCCGCGCCAGGGGGCCCCCAGGTTGGGTGTTGCGCGCCGGACTTAGGAAGACTTCCGACGGGTCG 144
         | .:.|||| || ||.||||||||| .||||.:||. ||| ||| .|| .|||.| ||.||||| ||
Db   616 TGTYTKKSGKGYGGCCHCGGSGSGGCKSTCGGSGSGGGGYGKSKCSGSGGCGSGGSGCCGGSGSG 675

Qy   145 CAACCTCGTGGAAGGCGCAGAAACCTATCCCCAAGGCTCGCCAGCCCGAGGGGCGCAGGGGCCTTGG 204
         :||| | ||||| |  |  ||.:. |||||||||||.|  | | ||||| || |||.||||||||.|
Db   676 CGGCSCSSSSGGCKCGSKSGGCSCCSBBSGGYCCBGGGSSSCKCSGSGGCSGSGGGCCGGSGGSG 735

Qy   205 GCTCAGCGCCCGGGTACCTTGGCCCCCTTATGGCAACGAGGGCATGGGGGTGGGCAGGGATGG 264
         | ||||| | ||||||.||||||||||| |||| | | ||||||.||||||||||||.||||| |
Db   736 CCCCTCCCCKBCGBGGGCSYGYCCGGSKCCBCTGKTCCCSSCYSTCKKSKCGCGTGTCK 795

Qy   265 CTCCTGTCACCCGGTGGGCCTTCCGGGCCCCACTGACCCCCGGGCCGGTGTCG 324
         |.| || .| ||||| | |.|||.|||| | || ||||||||||||| ||.|| |
Db   796 CKCCTGTGCKBYCKCYCTTCKYSYCKCBCYKCGCGCGGCSBKSCCGSCKCKTKKCY 855

Qy   325 CGTAATTTGGGTAAAGTCATCGATACCCTCACAATGCGCGCTTCGCGCTCTCATGGGGGTA 384
         :|.:|| :||.||.| |||| ||| ||||.||| | ||.|.||| | :||.||||| ||
Db   856 KKKKKKYBKRKYKKCCCKTYKCKCBCCKYYKCBBKCKCKCKCKCKCKKCBKKC 915

Qy   385 CATTCCGCCTCGTCGGCGGGCGCTTCCTTATGGGGGCGCCCCTTGGGGGCCGCATGGGGGTCTCCG 444
         | |||.|||.|| ||.||||.|| ||.||| ||||||||| | ||||| |.||||||||| |||
Db   916 BKCBKCBCKCKCBCKBSKCCCKBCCKYKCKCCKBKKBYKKCCBCBCKC 975

Qy   445 GTTCTGGAGGACGGCGGTGAAACTATGCAACAGGGAAATTTACCCGGTTGCTCTTTCTCTATC 504
         | .||||||||| |.|||.|| | |.|| | ||| | | |.|| ||||| | |||| ||
Db   976 KCCBCBCKKCKCKCBBKCBCKCBBKCBCKCKKCKCKCKCKCKKCKKCKKKYT 1035

Qy   505 TTCCTCTTTGGCGTTTGCTGTCCTGTTTGACCATTCCAGCTTC 545
         | :| :||||.|||.:||| | ||.|||  | :|||.|||.|:||
Db  1036 YTBKKKYKKKKKKKTBKKKBYKCKCCYKCBSCBKKCCKKBBKC 1076
```

RESULT 2

ORIGIN
```
                                                                /clone="BACR19D16"
                                                                /clone_lib="RPCI-98"
                                                                /note="end : TET3"

Query Match     6.7%;  Score 56.2;  DB 9;  Length 925;
Best Local Similarity 13.4%;  Pred. No. 0.0009;
Matches     48;  Conservative 167;  Mismatches 144;  Indels 0;  Gaps 0;

Qy   104 GGTTGGGTGTTGCGCGCCGGACTAGGAAGACTTCCGAGCGGGTCGCAACCTCGTGGAAGGCGAC 163
         |.:||.||:|||.|:||||||||||||||||||||||.||||||| |||||.|.:|||| || ||
Db   557 GSGYGKGCSSGSGGSBSCSCCSCKSSCSCCCHBCCCSCCCCCCSSYCCSSSBSSSSKCSSTSBS 616

Qy   164 AACCTATCCCCAAGGCTCGCCAGCCCGAGGGGCGCAGGGGCCTTGGGGGTCAGCGCCCGGGTTACCCTT 223
         |:||||||||| |||||.|||||.||| ||||||.|||||.||.||||||| | | ||||| ||| |.|
Db   617 CSCCCSSKSVCGTSCSSSSSSCSSSSSSSSTSSSSTSKSSSGSSSSSYTTSKSTS 676

Qy   224 GGCCCCTCTTATGGCAACGAGGGCATGGGGGTGGGCAGGATGGCTTCCTGTCACCCGTGGCT 283
         ::|||||.||  :||.| ||||||||||| |||||| |||||| |||.| ||||| || ||||
Db   677 ASGGSGSWSAGGGSGSSTGSTTSSSSSSSSTSSSSSVSSGSKSSTBSSGSSSGSSSSST 736

Qy   284 CCCGGCCCTAGTTGGGGGCCCCACTGACCCCCGGGCCGGTAGGTCGCGGTAATTGGGGTAAAGTCA 343
         ||||||.| ||| ||||||||| || ||||||||| |.||||.|| ||||.||||| |||||| ||
Db   737 SSBBSCTSTSSSSSSSYSSSSTCSCCCCSSSYSSSSTSGSSTSGSGSSSSSVGTSS 796

Qy   344 TCGATACCCTCACATGCGCGCTTCGCGCTCTCATGGGGGTACATTCCGCTTGCGTGGCGCT 403
         :|.|||||||.||.||| ||| |:||||.|.||||| ||.| | ||||| | || || ||
Db   797 SSDSTSTCCSCCCYMCTCCSTYBMHCYTSTSCGGSSSSGKGGVTKCGCGGCGGSSSTNGM 856

Qy   404 CCCTTAGGGCGCGCTGCCCAGGGCCCCTGCCGGTTCTGGAGGACGGCGGTG 462
         ::| ||||||||| | |.|| |||.|||| || |.||||||||| |:|| |
Db   857 BGTSSACSSSSSCSSSSSCVSSASSSSVSSGSSGVSSNSSSASKSSSSGSVSSG 915
```

RESULT 3

COMMENT   This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES   Location/Qualifiers
source   1..879
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="158C06"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG158BB03LP1-end : T7"

ORIGIN

Query Match   6.4%;   Score 53.4;   DB 9;   Length 879;
Best Local Similarity   41.3%;   Pred. No. 0.0049;
Matches 129;   Conservative   28;   Mismatches 155;   Indels   0;   Gaps   0;

```
Qy    14 GTAACACAAACCTCCGCCCACAGAGACGGTCAGGTTCCCGGCGSGTGGTCAGATCGTTGGTG   73
            ::  :  : :   :||||||||||||:||||||||||||||||:
Db   503 GYBBCCCCCCCCCCCCCCCCCCCCSGBSSVGGGMGGKSGGGGGGCCCCCCCCCCCCSGGGG  562

Qy    74 GAGTTTACCTGTTGCGCGCCAGGGGTGGGGTGTGCGCGGCGACTAGGAAGACTT  133
            ||||:  :  :   :   :   :   :
Db   563 GGGGSGGGGGGGCCCCCCCCCCCSCSSSCCGGGGGSGGGGGSSSSSCCCGCCCCC  622

Qy   134 CCGAGCGGTGCCAACCTCGTGGAAGGCGGACAACCTATCCCGGGCCTCGCCAGCCCGAGG  193
            ||||  :  : :  :   : :  :   :   :
Db   623 CCCCCGGGGCCTGGGGGCCGGGSGGGGGCSGGGGCCNNGGCCCCSSSSCCCCCCCCCC  682

Qy   194 GCAGGGGCCCTGGGGTTCAGCGCCGGGTACCCTTGGCCCCTCTATGGCAAC  253
            ||||:  :  :   :   :   :   :
Db   683 GGGGGCCCCGGGGSCCCGGGSGGSCCCCCCCGGSGGGCCCCCCCGGSGGGG  742

Qy   254 GGGGCAGGAGGATGGCTTCCTTGTCACCCCGGCGTTCCCGGGGTTGGGGGGCCCTGACCCCC  313
            ||||  :  :   :   :   :   :
Db   743 GGGGGCGGCCCCCCCCCCCCCCCCCCCCCCCCCSSCGSGSGGGGGGGGGGGGGG  802

Qy   314 GGGGCTAGGTCGC  325
            ||||:  :
Db   803 GCGGGCGGGGGGC  814
```

RESULT 4
CNS016UE
LOCUS   CNS016UE   1101 bp   DNA   linear   GSS 26-JUL-1999
DEFINITION   Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16G16 of DrosBAC library from Drosophila melanogaster (fruit
fly)., genomic survey sequence.
ACCESSION   AL107216
VERSION   AL107216.1   GI:5626210
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1   (bases 1 to 1101)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscape - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBeloBAC11.
FEATURES   Location/Qualifiers
source   1..1101

ORIGIN

Query Match   5.8%;   Score 48.6;   DB 9;   Length 1101;
Best Local Similarity   28.4%;   Pred. No. 0.092;
Matches 90;   Conservative   83;   Mismatches 144;   Indels   0;   Gaps   0;

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN16G16"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

```
Qy     1 AGAAAAACCAAAACGTAACACCAACCTTCCCGCCCCACAGGACGTCAGGTTCCCGGCGCGTGGT   60
            |||||||:   :   :   :||||:  :||::|:|  :::
Db   723 AAAAAAAAAAAAAAAAAAAAAAAAAAMCMCCCCYSSSGGGTCSSSGKGKSBGCCKCYYYYSK  782

Qy    61 CAGATCGTTGGTGTGGGAGTTTACCTGTTGCCGCGCCAGGGGGCCCCAGGGTTGGGTGTGCGCGGCG  120
            :   :   :   :   :   :   :   :
Db   783 SSKSSCGCSCKBGSGKWGGTTKTGTTGGGGSGGGSGGSCTRSGCYTSCCKTSCGCGCTSGT  842

Qy   121 ACTAGGAAGACTTCCGAGCGGTGCGCAAACCTCGTGGAAGGCGGACAACCTATCCCCAAGGCT  180
            :   :   :   :   :   :   :
Db   843 KTTGKGKTCCCCCTSYACCCCCHSSCCKBGYSTGGSGGGGSSGGGGGGSSCCSSCCSTGBGCSBSYY  902

Qy   181 CGCCAGCCCGAGGGCGAGGGGCCCTGGGGTTCAGCGCCGGGTACCCTTGGCCCCTCTATGGCAAC  240
            :   :   :   :   :   :   :
Db   903 BCCKGGKKGSGKKKGGYGTTTTGSCCSSCTYCCSCHWSSSWCYMYTKSCCCCCYTKYGCBCC  962

Qy   241 GAGGGCATGGGGGTGGGGCAGGAGGATGGCTTCCTGTTGTCACCCGCGGCCTAGTTCGGGG  300
            :   :   :   :   :   :   :
Db   963 BCSYGGCCTCYGCKYGBPGBGKGGKKCYGCGYCYCCTSGSSYCBBYYTTTCTSGKKSSSSC 1022

Qy   301 CCCACTGACCCCCGCG  317
            :   :   :||:|||:
Db  1023 GCYTCBSYGYSCTGGGG 1039
```

RESULT 5
CNS00BNG
LOCUS   CNS00BNG   1101 bp   DNA   linear   GSS 04-JUN-1999
DEFINITION   Drosophila melanogaster genome survey sequence TE73 end of BAC #
BACR24B13 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION   AL057398
VERSION   AL057398.1   GI:4937864
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1   (bases 1 to 1101)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y?; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
   source   Location/Qualifiers
               1..1101
              /organism="Drosophila melanogaster"
              /mol_type="genomic DNA"
              /clone="BACR24B13"
              /db_xref="taxon:7227"
              /clone_lib="RPCI-98"
              /note="end : TET3"

ORIGIN

Query Match   5.8%;   Score 48; DB 9; Length 1101;
Best Local Similarity   23.5%;   Pred. No. 0.13;
Matches 71; Conservative 94; Mismatches 137; Indels   0; Gaps   0;

Qy   95 GGGGGCCCCAGGTTGGGGTGTGCCGCGGACTAGGAGAGAGACTTCGAGGGGTCGCAACCTCGTG 154
               |:|:|
Db   701 GGWGMWGGMGRGKGMGKGMGNGMGNGMGKTKSGKMCGGGCGBKCCGGKNKTKCCKCKCKK 760

Qy   155 GAAGGGCGACAACCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGGCTGGGGTCAGCCCG 214
               |:|:|
Db   761 CCGKNCKHCBCKCCCKNKKKTCCKCKBCCGKGKGCCKKKYGKKNGKCCTTTTTTCTNTYTC 820

Qy   215 GGTACCCTTGGCCCCTCTATGGCAACAGAGGGCCATGGGGTGGCAGGAGTCGCTCCTGTCAC 274
               |:|:|
Db   821 CYTTTCTTTCTTTCYCTTTTTTTKKKKMNMNKRYGKKKKKKMKKKKNCKKCGYCCKKKYGKN 880

Qy   275 CCCGTGGCTCCCGGGCCTTAGTTGGGGCCCCCACTGACCCCCGTAGGTCGGGTAATTTGG 334
               |:|:|
Db   881 YCYBKKBCCCCCKCCCCKCCTCBTYKYTGBNBGKKKKKKKTCKSKCSKBNBGKKKKTCKKKKKKCCB 940

Qy   335 GTAAAGTCATCGATACCCTCACATGCGGCGCCTTGCGCTCTCATGGGGTACATTTCCGTC 394
               |:|:|
Db   941 KCKYGKNCCBKKKKKCCCKKCCCCGSCKCKCKCGKKKKKKCKCMCCKCCCCCKC 1000

Qy   395 GT 396
               |:|
Db   1001 GK 1002

RESULT 6
LOCUS   CNS010BS
DEFINITION CNS010BS   1036 bp   DNA   linear   GSS 26-JUL-1999
          Drosophila melanogaster genome survey sequence SP6 end of BAC
          BACN03H11 of DrosBAC library from Drosophila melanogaster (fruit
          fly), genomic survey sequence.
ACCESSION   AL098770
VERSION   AL098770.1 GI:5610381
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1036)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
          collaboration with the European Drosophila melanogaster BAC
          project (Dros BAC) was made by Alain Billaud at CEPH (Centre
          d'Etude du Polymorphisme Humain) with funding provided by a MRC
          http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster
          library (EDGP) - http://www.edgp.ebi.ac.uk -.
          project grant. The DNA was prepared from embryos by Alain Bucheton
          and Genevieve Payan. It has been constructed in the vector
          pBeloBAC11.

FEATURES
   source   Location/Qualifiers
          1..1036
          /organism="Drosophila melanogaster"
          /mol_type="genomic DNA"
          /db_xref="taxon:7227"
          /clone="BACN03H11"

ORIGIN
          /clone_lib="DrosBAC"
          /plasmid="pBeloBAC11"
          /note="end : SP6"

Query Match   5.7%;   Score 47.8; DB 9; Length 1036;
Best Local Similarity   26.3%;   Pred. No. 0.15;
Matches 114; Conservative 117; Mismatches 199; Indels   3; Gaps   1;

Qy   1 AGAAAAACCAAAACGTAAACACCAACCTCCGCCCACAGAGGACGTTCAGGTTCCGGGGGGGGGGT 60
               ||||| |||| ||:|
Db   593 ATANAAAAAAAATAWACAAAYAAGCCCCSAAACCCACCACMCJAACKACCCSSSAVSMSGA 652

Qy   61 CAGATCGTTGGTGGAGTTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGTGCGCGG 120
               :: ||::| |::::::| | ||||:
Db   653 SARAAAAAAVSAAVAGMGAAAAAAAGASCYGGAGSGRRSSSSCRGCCSVCSYCKGGCCCY 712

Qy   121 ACTAGGAAGACTTCCGAGCGGGGTCGCCAACCTCGTGGGAAGGGCGACAACCTATCCCCAAGGCT 180
               ||| ||| | : ::||:| ||:::| |:
Db   713 GAAAGAAAGACGAGSAAV--GCGSAAGCAGSSAGCACAWMGASCSAASARAVCGGCCGSCGS 769

Qy   181 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGTACCCTTGGCCCCTTCTATGGCAAC 240
               ||||:| | ||:|| |||:|||:| | |||:|:
Db   770 CVSCASCSCGCCGCCSGSGMGSCGGAMASSSCCSSSCASWAAGASASGSGGSCGCSCGASSCVSSG 829

Qy   241 GAGGGCATGGGGTGGGCAGGAGTCGCTCCTGTCACCCCGTGGGCTCCCGGGCCTAGTTGGGGC 300
               || || | ||| | ||:| |:||| | | ||:|
Db   830 SCGCGMGCVSCGACSCCSSCGGCGCGSSSGMGSSSASSGSGSSAASGSGGGGCGCSCGASSGSSGSS 889

Qy   301 CCCACTGACCCCCGCGGGTAGGTCGGGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360
               |:|| |||:|:| | |:| ||:|||:
Db   890 GCGASRMASGSGSGGGSGVAARGCGSARAGCRRAGSSGCGCACNSCASMGCKCGMAAAA 949

Qy   361 GGCTTCGCCGACTTCTCATGGGGGTACATTTCCGTCGTCGGCGCCCCTTPAGGGGGGGCTGCC 420
               ||||| | ||:| | |||:| |:|| | ||:|
Db   950 SVSSGMGCAAAAAAARASGCGSSSSCCSGGSGGSGVGVSBSTCGSARVARGSGGCSAGCGSSS 1009

Qy   421 AGGGCCCTGGGCGC 433
               |:|||:|:| ::|
Db   1010 GGGSVSVSSVGSGSS 1022

RESULT 7
LOCUS   AG126333/c
DEFINITION AG126333/c   1030 bp   DNA   linear   GSS 04-NOV-2001
          Pan troglodytes DNA, clone: PTB-136N19.F, genomic survey sequence.
ACCESSION   AG126333
VERSION   AG126333.1 GI:16655498
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   BAC end sequences of Library PTB
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1030)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimpbee@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the R&D process and may have higher chance of
          clone tracking errors.
          PRIMERS
          Sequencing: -21M13
          LIBRARY

RESULT 8

LOCUS   CNS0091P/c
DEFINITION CNS0091P   925 bp   DNA   linear   GSS 03-JUN-1999
            Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION   AL053013.1   GI:4934461
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and

FEATURES   Location/Qualifiers
     source   1..925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

Query Match   5.5%; Score 45.8; DB 9; Length 925;
Best Local Similarity   12.1%; Pred. No. 0.49;
Matches   43; Conservative 163; Mismatches 149; Indels 0; Gaps 0;

Qy   88 CCGCGCAGGGGCCCCCAGGTTGGGTGTGGCGGCGACTAGGAGAACTTCCGACGGTCGCAA 147
              :::::: :|:: ::::::::: ::: :|::: : ::: :||::::::::: :::::::::::
Db   924 SBSCCSCSCSSCSSSBSCCSSSMSTSSSNSSBCSCSCSSSBSSSSTSSMSSSSGSSSSS 865

Qy   148 CCTCGTGGAAGGCGACAACCTTATCCCCAAGGCTCGCCACGCCCGAGGGCAGGGCCTGGGCT 207
              :: ::::::::::::::::::::: :::::: ::::::: :||:::::::::::::::::::::
Db   864 SGTSSACVKCNASSSCGCCCGCGMARCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSG 805

Qy   208 CAGCCCGGGTACCCTTGGCCCCTCTATGGCAACGAGGGCATGGGTGGGGGAGGATGGGCTC 267
              |: :::::::::SSACBSSSSSCSASCWSASSSSSASSSRSRSGGGGAGGSGASSSRSSSSSSA 745
Db   804 GASASHSSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGGAGGSGASSSRSSSSSSA 745

Qy   268 CTGTCACCCCGTGGCTCCCCGGCCCTAGTTGGGGGCCCCACTGACCCCCGGGCGTAGGTGCGT 327
              :::::::::::: :::::::::: ::::::::::::::::::: :::::::::::: :::::::
Db   744 SAGSVVSSASSSSCCSSSVSCSSSVVASSMCSSSBSSSSASASSSSSSASCASCSCCCCT 685

Qy   328 AATTTGGGTAAAAGTCATCGATACCCTCACATGCCGACTTCGCCGACTTCATGGGTACAT 387
              :: :::
Db   684 SWSCSCSTSASMSAARSSSSSSSCSSSSMSASSSASASSSSSSSSGGSACGBS 625

Qy   388 TCCGGTCGTCGGGCGTCCCTTAGGGGGCGTCGCCAGGGCCCCTGGCCCATGGGGGTCC 442
              :::: ::::::::::::::: :::: :::::::::::::: :::: :::::::::::::::
Db   624 MSSGGGSGSVVSASSGMSSSVSSSGGRSSSGSGGGGGVGGSGSSSGSGGSGSY 570

REFERENCE
  AUTHORS
  TITLE   Aaron Mammoser in Pieter de Jong's laboratory in the Department of
  JOURNAL   Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            PI and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/droosophila_bac.htm.

RESULT 9

LOCUS   CNS0052P/c
DEFINITION CNS0052P   844 bp   DNA   linear   GSS 03-JUN-1999
            Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION   AL056652.1   GI:4932342
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila

FEATURES   Location/Qualifiers
     source   1..1030
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /sex="male"
                     /clone="PTB-136N19.F"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

Query Match   5.7%; Score 47.6; DB 9; Length 1030;
Best Local Similarity   34.3%; Pred. No. 0.17;
Matches 134; Conservative   0; Mismatches 257; Indels   0; Gaps 0;

Qy   25 CTCCGCCCACAGGACACGTCAGGTTCCCGGCGGTGGTCAGATCGTTGGGTGGGAGTTTACTTG 84
              |||||||| : ::::: : || : : : : : ::::::::::::::::::: ::
Db   480 CGGCCACCACCCGCCGCGGGNNNNNNNGGCGGGGGCCGCGGCCGCNNCCGNNGGCGGG 421

Qy   85 TTGGCGCCAGGGGCCCCCAGGTTGTGCGGCCGGACTAGGGAAGACTTCCGACGGGTCG 144
              |||| ::::::: :::::::::: :::::::::::: :::::::::::::::::::::::::
Db   420 GGGCGGGGGGGGGGGCCCCGCGGGGGGGGGGGGCCCCGCGGGCCGCGCCGCGCCGCCGG 361

Qy   145 CAACCTCGTGGGAAGGCGACAACCTTATCCCCAAGGCTCGCCAGCCCGAGGGCCAGGGCCCTCG 204
              | :: ::: :: ::: ::::::: ::::::: ::::::: :||:::::::::::::::::::::
Db   360 GGCCCCCCGGNNGGGNGGGCCCAACAAAAGCGCCCGGGGCGCGGGGGGCGGGGGGGCGGG 301

Qy   205 GCTCAGCCCCGGGTACCCTTGGCCCCTCTATGGCAACGAGGGCATGGGTGGGGTGGGGACCATG 264
              :::::::::::::::::: ::: : :::: :: ::: : :: :::::::::::::
Db   300 GGGGGGNNGGGGGGNNGGGGGGNNGCCCCGNNGGGNNGGGGGGGGGNNGNGNGGGGNG 241

Qy   265 CTCCTGTCACCCCGTGGCCCTAGTTGGGGGGCCCCACTGACCCCCGGGGGCGTAGGTGCG 324
              ||||
Db   240 NNNNNNNGNGNGNNNNNNGGGNNNCNNNNNNNNNCCCCCNNCNCNNNNNNNNNNNN 181

Qy   325 CGTAATTTCGGTTAAAAGTCATCGATACCCTCACATGCCGGCTTCGCCGACTTCTCATGGGGTA 384
              ||:
Db   180 NGNGCNCCCNNNNCCNNNNCACCTTNNNNNNNNNGNNGGACNNCGGCCGCCAANNNCCCNN 121

Qy   385 CATTCCGGTCGTCGGGCGTCCCTTAGGGGGCG 415
              |:|
Db   120 CNNCCNCNNNNNNNNCNCCNNNNNNCCTTNGGGGG 90

COMMENT   melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES   Location/Qualifiers
        source   1..844
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11P16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

Query Match   5.4%;   Score 44.8;   DB 9;   Length 844;
Best Local Similarity 20.2%;   Pred. No. 0.89;
Matches 71;   Conservative 128;   Mismatches 153;   Indels   0;   Gaps   0;

Qy   179 CTCGCCAGCCCGAGGCAGGGCCCTGGGCTTCAGCCCGCGGTACCCTTGGCCCCTTATGGCA 238
              ||||| ||| ::::::: |||| ::::::: |||| ::::: :||||||||||
Db   644 CCCCYYCSCCBVBCNYSCSYYSYSBCSGCSCSSSSSSCCSSGGSSBYSCCCBCCCBGYCGCY 585

Qy   239 ACGAGGGGCCATGGGGGGTGGGCAGGATNGGCCTCCTGTCACCCCGTGGCTCCCGGTTCGGG 298
              :::::|||:::: |||| ::::: |||| ::::: |||| ::::: ||||
Db   584 SBVYSBSBBSSTBSSVSGBSSSTSSSGTGCKCCBSSSSSBBCCSSSCCCYCSSYCTYCYSSCC 525

Qy   299 GCCCCACTGACCCCCCGGGGGTAGGTCGCGTAAATTGGGTAAAGTCATCGATACCCTCACAT 358
              :::::|||:::::::: |||| ::::: ||||:::: |||| :::::
Db   524 BYCCBSCGCBTBSYBSCSSGBTGBSSBSGGKSGGKGGGBKSGGKGYCYCBYTYCCSBS 465

Qy   359 GCGGCTTCGCCGACTCTCATGGGGGTACATTCCGCTCGTCGGGGSTKBYSYSBTTBBYTBBTT 418
              ::::: |||| ::::: |||| ::::: |||| ::::: |||| :::::
Db   464 SBCCSSSSSSTSKCCSGGSSSTKCGSSBBGGGGSSKISGSSCBSGSSSSTSGSSGSSKKYS 405

Qy   419 CCAGGGCCCTGGCGCATGGGGTTCTCGGGTTTGAGGACGGGGCGTGAACTATGCAACAGGGA 478
              ::::: |||| ::::: |||| ::::: |||| ::::: ||||
Db   419 YSTSSSSSGCGKGKNTTTGCVGSTCTTTKSSCTBSSGGTGSSBSTSTTTGTTTTTTTTTG 345

Qy   479 ATTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGCTTTGCTCGTCCTGTTT 530
              ||| |||| | ||||||||||||||||||||||||| ||||| |||||
Db   344 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTGTTT 293

RESULT 10
LOCUS   CNS060ON   910 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065629
VERSION   AL065629.1 GI:4944698
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 910)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information

---

            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES   Location/Qualifiers
        source   1..910
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14J21"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

Query Match   5.4%;   Score 44.8;   DB 9;   Length 910;
Best Local Similarity 14.7%;   Pred. No. 0.9;
Matches 43;   Conservative 146;   Mismatches 101;   Indels   2;   Gaps   1;

Qy   171 CCCCAAGGCTTCGCCGAGGGCAGGGCCTGGGGCTCAGCCCGCGGTACCCTTGCGCCCT 230
              |||::::::: |||| ::::: |||| ::::: |||| ::::: ||||
Db   556 CSSSSSSSSSSSSCCGGSSSBBGCBSSCCGCGCSCCSSCSYSSSSTGSTGSGCCCGSSSSSSSCCST 615

Qy   231 CTATGGCAACGAGGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGCC 290
              ::::: |||| ::::: |||| ::::: |||| ::::: |||| :::::
Db   616 STSSSSTSGCYGYSSSGSSBSSSCSSTBTSSSTTBTKBTSYSYSYTSTTTSTTSTBSBT 675

Qy   291 TAGTTGGGGCCCCCACTGACCCCCGGGGTAG--GTCGCGTAAATTGGGTAAAGTCATCGAT 348
              ::::: |||| ::::: |||| :::::   ::::: |||| :::::
Db   676 SYGSBBKSTSSTSSSTSTSSSSTSSSSSCTSTSSSCTSTBSYSBTTBYTBBTTSTSTBBT 735

Qy   349 ACCCTCACATGCGGCCTTCGCCGACTCTCATGGGGTACACATTCCGCTCGTCGGGCCTT 408
              ::::: |||| ::::: |||| ::::: |||| ::::: :::::
Db   736 TSTBYTBBTTBYTKSTTSMTSTYYTBBBSTSKBSTBTSTBKSTGTKTBSBTTSBTTSCTSSSSS 795

Qy   409 AGGGGCGGCCTGCCAGGGCCCCTGCGCGCAGGGCGTCCGGGTTCTGGAGGACGGCG 460
              ::::: ::::: |||| ::::: |||| ::::: |||||
Db   796 BSTSYSYSYSTSCBSSBSSBSSTSYSBCTSTSTSTSSBBBSSBSSSSSCCCGTSBTSS 847

RESULT 11
LOCUS   CO141650   439 bp   mRNA   linear   EST 17-JUN-2004
DEFINITION EST836321 Aspergillus flavus Normalized cDNA Expression Library
            Aspergillus flavus cDNA clone NAFEP03 5' end, mRNA sequence.
ACCESSION   CO141650
VERSION   CO141650.1 GI:4889293
KEYWORDS   EST.
SOURCE   Aspergillus flavus
  ORGANISM Aspergillus flavus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1 (bases 1 to 439)
AUTHORS   Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.
TITLE   Aspergillus flavus expressed sequence tags for identification of
            genes with putative roles in aflatoxin contamination of crops
JOURNAL   FEMS Microbiol. Lett. (2004) In press
COMMENT   Contact: Yu J
            Food and Feed Safety Research Unit
            USDA/ARS, Southern Regional Research Center
            1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
            Tel: 504 286 4405
            Fax: 504 286 4419
            Email: jiuyu@srrc.ars.usda.gov
            Contact Dr. Yu at USDA/ARS SRRC (jiuyu@srrc.ars.usda.gov) for clone
            information
            PCR Primers

Left column

```
FORWARD: M13F
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES             Location/Qualifiers
source               1..439
                     /organism="Aspergillus flavus"
                     /mol_type="mRNA"
                     /strain="NRRL 3357"
                     /db_xref="taxon:5059"
                     /clone="NAFF03"
                     /sex="asexual mycelia"
                     /cell_type="mycelia"
                     /dev_stage="developmental stages from 18 to 96 hours"
                     /lab_host="E. coli DH10B T1 resistant cells"
                     /clone_lib="Aspergillus flavus Normalized cDNA Expression
                     Library"
                     /note="Vector: pBlueScript (SK+) (Stratagene), antibiotic
                     selection marker: Carbenicillin; Site_1: NotI, at the 5
                     prime end; Site_2: EcoRI, at the 3 prime end; This
                     normalized cDNA expression library was constructed using a
                     mixture of mycelial cells grown under eight different
                     medium conditions and harvested at 5 time points (18, 24,
                     48, 72, 96 hours). The poly-A sequence was trimmed off
                     before ligating to vector."
ORIGIN
```

```
Query Match       5.3%;    Score 44.4;    DB 7;    Length 439;
Best Local Similarity   49.2%;    Pred. No. 1;
Matches  117;  Conservative  0;  Mismatches  121;  Indels  0;  Gaps  0;

Qy  579 CCAAGTCACGAAACGATTGCTTCCAACTCAACCATCCTGTACGAGACGCAGCGGACACATGATCAT 638
            |||| || ||  |||  |   || || ||| |  |    || || ||||  ||||||||||
Db   32 CAAGGCGCCGGCCCCCCAAGCACCCTCTAAGATCCCAGATGTGTCTTCCGAAGCCAACAA  91

Qy  639 GCACACCCCCGGGGTGTGTGCCCCTGTTGTCCGGGAGGGGTAATTCCTCCCGCCTGCTGGGGTAGC 698
            || | ||||  |  ||  | |  || | |  |  |   | |||||  ||| || |||| ||
Db   92 GCTCGCCCCCAAACGGCGAGTTCGGGTCGATCGGGCTACTGCTTGCGGGTGGAAA 151

Qy  699 GCTCACTCCCACGCTCGCGCGGCCCAAGGACGCCCAGCAATCCCCACTGCCACAATACGACGCCA 758
            ||||| | ||  || ||| || ||||| || |  || | |||| | |||| || ||||||
Db  152 GATCACCGCCCCTGGCCTCGGGCAAGGACAACAAGATCTTCAAGGCGCCGCCCGTGCAGTGCCA 211

Qy  759 CGTCGGATTTTGCTCGTTGGGGGGGCGGCTGCCTTCTCCGTCGCTTATGTACGTTGGGGGGATCTC 816
            ||  ||   |||| | |      |  ||  | |||  || ||| |||  ||||
Db  212 CCCCGCTATGCTTGCCCCAGAAGAGTCGGTCTCAATCCCCATGGCCGTGCTC 269
```

```
RESULT 12
CO137934
LOCUS       CO137934      450 bp    mRNA    linear    EST 17-JUN-2004
DEFINITION  EST832605 Aspergillus flavus Normalized cDNA Expression Library
            Aspergillus flavus cDNA clone NAFCN29 5', end, mRNA sequence.
ACCESSION   CO137934
VERSION     CO137934.1  GI:48886912
KEYWORDS    EST.
SOURCE      Aspergillus flavus
  ORGANISM  Aspergillus flavus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.
  TITLE     Aspergillus flavus expressed sequence tags for identification of
            genes with putative roles in aflatoxin contamination of crops
  JOURNAL   FEMS Microbiol. Lett. (2004) In press
COMMENT     Contact: Yu J
            Food and Feed Safety Research Unit
            USDA/ARS, Southern Regional Research Center
            1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
            Tel: 504 286 4405
            Fax: 504 286 4419
            Email: jiuyu@srrc.ars.usda.gov for clone
            Contact Dr. Yu at USDA/ARS SRRC (jiuyu@srrc.ars.usda.gov) for clone
```

Right column

```
information
PCR PRimers
FORWARD: M13F
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES             Location/Qualifiers
source               1..450
                     /organism="Aspergillus flavus"
                     /mol_type="mRNA"
                     /strain="NRRL 3357"
                     /db_xref="taxon:5059"
                     /clone="NAFCN29"
                     /sex="asexual mycelia"
                     /cell_type="mycelia"
                     /dev_stage="developmental stages from 18 to 96 hours"
                     /lab_host="E. coli DH10B T1 resistant cells"
                     /clone_lib="Aspergillus flavus Normalized cDNA Expression
                     Library"
                     /note="Vector: pBlueScript (SK+) (Stratagene), antibiotic
                     selection marker: Carbenicillin; Site_1: NotI, at the 5
                     prime end; Site_2: EcoRI, at the 3 prime end; This
                     normalized cDNA expression library was constructed using a
                     mixture of mycelial cells grown under eight different
                     medium conditions and harvested at 5 time points (18, 24,
                     48, 72, 96 hours). The poly-A sequence was trimmed off
                     before ligating to vector."
ORIGIN
```

```
Query Match       5.3%;    Score 44.4;    DB 7;    Length 450;
Best Local Similarity   49.2%;    Pred. No. 1;
Matches  117;  Conservative  0;  Mismatches  121;  Indels  0;  Gaps  0;

Qy  579 CCAATGTCACGAAACGATTGCTTCCAACTCAACCATCGTGTACGAGACGACGCAGCGGACACATGATCAT 638
            |||| || ||  |||  |   || || ||| |  |    ||  || ||||  ||||||||||
Db   43 CAAGGCGCCGGCCCCCCAAGCACCCTCTAAGATCCCAGATGTGTCTTCCGAAGCCAACAA 102

Qy  639 GCACACCCCCGGGGTGTGTGCCCCTGTTGTCCGGGAGGGGTAATTCCTCCCGCCTGCTGGGGTAGC 698
            || | ||||  |  ||  | |  || | |  |  |   | |||||  ||| || |||| ||
Db  103 GCTCGCCCCCAAACGGCGAGTTCGGGTCGATCGGGCTACTGCTTGCGGGTGGAAA 162

Qy  699 GCTCACTCCCACGCTCGCGCGGCCCAAGGACGCCCAGCAATCCCCACTGCCACAATACGACGCCA 758
            ||||| | ||  || ||| || ||||| || |  || | |||| | |||| || ||||||
Db  163 GATCACCGCCCCTGGCCTCGGGCAAGGACAACAAGATCTTCAAGGCGCCGCCCGTGCAGTGCCA 222

Qy  759 CGTCGGATTTTGCTCGTTGGGGGGGCGGCTGCCTTCTCCGTCGCTTATGTACGTTGGGGGGATCTC 816
            ||  ||   |||| | |      |  ||  | |||  || ||| |||  ||||
Db  223 CCCCGCTATGCTTGCCCCAGAAGAGTCGGTCTCAATCCCCATGGCCGTGCTC 280
```

```
RESULT 13
CO138520
LOCUS       CO138520      743 bp    mRNA    linear    EST 17-JUN-2004
DEFINITION  EST833191 Aspergillus flavus Normalized cDNA Expression Library
            Aspergillus flavus cDNA clone NAFCU57 5', end, mRNA sequence.
ACCESSION   CO138520
VERSION     CO138520.1  GI:48887498
KEYWORDS    EST.
SOURCE      Aspergillus flavus
  ORGANISM  Aspergillus flavus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.
  TITLE     Aspergillus flavus expressed sequence tags for identification of
            genes with putative roles in aflatoxin contamination of crops
  JOURNAL   FEMS Microbiol. Lett. (2004) In press
COMMENT     Contact: Yu J
            Food and Feed Safety Research Unit
            USDA/ARS, Southern Regional Research Center
            1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
            Tel: 504 286 4405
            Fax: 504 286 4419
```

RESULT 14 (left column)

Email: jiuyu@srrc.ars.usda.gov
Contact Dr. Yu at USDA/ARS SRRC (jiuyu@srrc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13F
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES   Location/Qualifiers
     source   1..743
                  /organism="Aspergillus flavus"
                  /mol_type="mRNA"
                  /strain="NRRL 3357"
                  /clone="NAFCU57"
                  /db_xref="taxon:5059"
                  /sex="asexual mycelia"
                  /cell_type="mycelia"
                  /dev_stage="developmental stages from 18 to 96 hours"
                  /lab_host="E. coli DH10B T1 resistant cells"
                  /clone_lib="Aspergillus flavus Normalized cDNA Expression
                  Library"
                  /note="Vector: pBlueScript (SK+) (Stratagene), antibiotic
                  selection marker: Carbenicillin; Site_1: NotI, at the 5
                  prime end; Site_2: EcoRI, at the 3 prime end; This
                  normalized cDNA expression library was constructed using a
                  mixture of mycelial cells grown under eight different
                  medium conditions and harvested at 5 time points (18, 24,
                  48, 72, 96 hours). The poly-A sequence was trimmed off
                  before ligating to vector."

ORIGIN

Query Match   5.3%; Score 44.4; DB 7; Length 743;
Best Local Similarity 49.2%; Pred. No. 1.1;
Matches 117; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy    579 CCATGTCACGAACGAATTGCTTCAACTCAAGCATCGTGTGTTACGAGACACGGGACATGATCAT 638
              ||||| |||| | ||| ||||| | ||||| | |||| ||| |||||| | ||  |||
Db    424 CAAGGCGCCGACGCCCCCAAGCACCCTTCTTAAGATCCCAGATGTTGTTCTCCGAAGCCAACAA 483

Qy    639 GGCACACCCCCGGGGTGTGTCCCCTGTGTCCGGGGAGGGTAATTCCTCCGCTGCTTGGGGTAGC 698
Db    484 GCTCGCCCCAACGGGCAGAGTTCGGGTCCTCGGGTGCTGATCCTGGCCTACTGCTTGGGGTGAAA 543

Qy    699 GCTCACTCCCACGCTCGCGGCCAAGGAGGAGGGCCAGCGATCCCACTGGGCGACAATACGACGGCA 758
Db    544 GATCACCCCCCTCGGGCCTCGGGCAAGGACAACAAGAGATCTTCAAGGCCCGCGTGCAGTGCCA 603

Qy    759 CGTCGGATTTGCTTCCGTTGGGGGCGGCGGCTGCCTTCTCGTCGCCTATGTACGTGGGGGGATCTC 816
Db    604 CCCCGCTATGCTTGCCCCAGAAGAATGCCAAGTCGGTTCTCAATCCCCATGGCCGTGCTC 661
```

LOCUS   CO143444   548 bp   mRNA   linear   EST 17-JUN-2004
DEFINITION EST838115 Aspergillus flavus Normalized cDNA Expression Library
            Aspergillus flavus cDNA clone NAFFB15 5' end, mRNA sequence.
ACCESSION   CO143444
VERSION   CO143444.1 GI:4897445
KEYWORDS   EST.
SOURCE   Aspergillus flavus
  ORGANISM Aspergillus flavus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1 (bases 1 to 548)
  AUTHORS   Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.
  TITLE   Aspergillus flavus expressed sequence tags for identification of
            genes with putative roles in aflatoxin contamination of crops
  JOURNAL   FEMS Microbiol. Lett. (2004) In press
COMMENT   Contact: Yu J
            Food and Feed Safety Research Unit
            USDA/ARS, Southern Regional Research Center
            1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA

RESULT 15 (right column)

Tel: 504 286 4405
Fax: 504 286 4419
Email: jiuyu@srrc.ars.usda.gov
Contact Dr. Yu at USDA/ARS SRRC (jiuyu@srrc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13F
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES   Location/Qualifiers
     source   1..548
                  /organism="Aspergillus flavus"
                  /mol_type="mRNA"
                  /strain="NRRL 3357"
                  /clone="NAFFB15"
                  /db_xref="taxon:5059"
                  /sex="asexual mycelia"
                  /cell_type="mycelia"
                  /dev_stage="developmental stages from 18 to 96 hours"
                  /lab_host="E. coli DH10B T1 resistant cells"
                  /clone_lib="Aspergillus flavus Normalized cDNA Expression
                  Library"
                  /note="Vector: pBlueScript (SK+) (Stratagene), antibiotic
                  selection marker: Carbenicillin; Site_1: NotI, at the 5
                  prime end; Site_2: EcoRI, at the 3 prime end; This
                  normalized cDNA expression library was constructed using a
                  mixture of mycelial cells grown under eight different
                  medium conditions and harvested at 5 time points (18, 24,
                  48, 72, 96 hours). The poly-A sequence was trimmed off
                  before ligating to vector."

ORIGIN

Query Match   5.2%; Score 43.6; DB 7; Length 548;
Best Local Similarity 49.1%; Pred. No. 1.7;
Matches 115; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```
Qy    583 GTCACGAACGAATTGCTTCAACTCAAGCATCGTGTGTTACGAGACACGGGACATGATCATGCAC 642
              ||||| |||| | ||| ||||| | ||||| | |||| ||| |||||| | ||  |||
Db      1 GCCGCGCGCCCCAAGCACCCTTCTTAAGATCCCAGATGTTGTTCTCCGAAGCCAACAAGCTC 60

Qy    643 ACCCCCGGGGTGTGTCCCCTGTGTCCGGGGAGGGTAATTCCTCCGCTGCTTGGGGTAGCGCTC 702
Db     61 GCCCCCAACGGGCAGAGTTCGGGTCCTCGGGTGCTGATCCTGGCCTACTGCTTGGGGTGAAAGATC 120

Qy    703 ACTCCCACGCTCGCGGCCAAGGAGGAGGGCCAGCGATCCCACTGGGCGACAATACGACGGCACGTC 762
Db    121 ACCGCCCTGGGCCTCGGGCAAGGACAACAAGAGATCTTCAAGGCCCGCGTGCAGTGCCACCCC 180

Qy    763 GATTTGCTCGTTGGGGGCGGCGGCTGCCTTCTCGTCGCCTATGTACGTGGGGGGATCTC 816
Db    181 GCTATGCTTGCCCCAGAAGAATGCCAAGTCGGTTCTCAATCCCCATGGCCGTGCTC 234
```

LOCUS   CD109759   899 bp   mRNA   linear   EST 15-MAY-2003
DEFINITION AGENCOURT_13988238 NIH_MGC_147 Homo sapiens cDNA clone
            IMAGE:30347093 5', mRNA sequence.
ACCESSION   CD109759
VERSION   CD109759.1 GI:30753968
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 899)
  AUTHORS   Strausberg,R.
  TITLE   NIH-MGC http://mgc.nci.nih.gov/.
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson

cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I:M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM387 row: k column: 06
High quality sequence stop: 265.

FEATURES   Location/Qualifiers
    source   1..899
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:30347093"
                        /tissue_type="Human Placenta"
                        /lab_host="DH10B TonA"
                        /clone_lib="NIH_MGC_147"
                        /note="Organ: Placenta; Vector: pBluescriptR; Site 1:
                        all-XhoI; Site 2: BamH; Oligo-dT primed using primer
                        5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                        insert size 2.3 kb and normalized to ROT 5. This is a
                        primary library enriched for full-length clones and
                        constructed using the Cap-trapper method (Carninci, in
                        preparation). Library constructed by M. Brownstein
                        (NIMH/NHGRI, National Institutes of Health). Note: This is
                        a NIH_MGC library."

ORIGIN

Query Match   5.2%;   Score 43.2;   DB 6;   Length 899;
Best Local Similarity   47.1%;   Pred. No. 2.4;
Matches 132;   Conservative   0;   Mismatches 148;   Indels   0;   Gaps   0;

Qy   174 CAAGGCTTCGCCAGCCCGAGGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTA 233
               ||||||||||||||||| |||||||||||||||||||||| |||||||||||||||||||
Db   24 CCAGGCCCTGCCCCCCGCTGGGCCCCCCGGGAGGAGCGGAGCGCGGCTCACCCTTCGCCCCGGGGGC 83

Qy   234 TGGCAACGAGGGGCATGGGGGTGGGCGAGGATGGCTTCCTGTCACCCCGTGGCTTCCCGGCCTAG 293
               ||||||||||||| | |||||||||||||||||||||| |||||||||||| |||||||||||
Db   84 TGGGAGGGCGGCGCGGAGTCGGGGTGCACGGCGCACCCCCTGCCCCGCCCCTTCGGGCCCCTTCCCCG 143

Qy   294 TTGGGGGCCCCACTGACCCCCGGGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCT 353
               |||||||||||| ||||||||||||||||||||| | ||| |||||||||||||||||||
Db   144 CGGGCGGGGTGCAGCTACCCCTGCAGGGCCTTCCCCTAGCTAGAAGGGAGCGGGGAGGGGGCTC 203

Qy   354 CACATGCGCGGCCTTCGCCGACTCTCATGGGGTTACATTCCCGCTCGTCGGGGGCTCCCTTAGGGG 413
               ||||| |||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   204 CGGGGCGCCCGCGCAGCAGACCCTGCTTCCGGCCCGCGCGGCCCTCGCCCGCGTGTCCTCCCGGGGAGCGG 263

Qy   414 CGCTGCCCAGGGCCCTGGCCCATGGCGTCCCGGGTTCTGGGAG 453
               ||| ||||||||||||| ||||| |||||||| ||||
Db   264 CAGCCCTAGCCCCGGGCGCGGGAGGGGCTGGGGGGTTCCTCGAG 303

Search completed: February 21, 2005, 04:05:44
Job time : 3567 secs

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: February 21, 2005, 01:19:35 ; Search time 194 seconds
(without alignments)
7034.303 Million cell updates/sec

Title: US-09-664-363-5
Perfect score: 834
Sequence: 1 AGAAAAACCAAACGTAACAC........TCTGCGGATCTGTTTCCCG 834
Scoring table: IDENTITY NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Issued Patents NA:*
   1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
   2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
   3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
   4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
   5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
   6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	834	100.0	834	3	US-08-191-160-5	Sequence 5, Appl
2	802.6	96.2	2116	3	US-08-191-160-21	Sequence 21, Appl
3	737	88.4	2433	3	US-08-612-973-49	Sequence 49, Appl
4	737	88.4	2433	3	US-08-927-597-49	Sequence 49, Appl
5	733.8	88.0	9413	4	US-09-827-688-6	Sequence 6, Appli
6	730.6	87.6	1539	2	US-08-470-426B-17	Sequence 17, Appl
7	730.6	87.6	1863	3	US-08-470-426B-14	Sequence 14, Appl
8	730.6	87.6	9595	3	US-09-014-416-4	Sequence 4, Appli
9	730.6	87.6	9599	3	US-09-014-416-6	Sequence 6, Appli
10	730.6	87.6	11076	4	US-09-539-601-1	Sequence 1, Appli
11	730.6	87.6	11076	4	US-09-539-601-19	Sequence 19, Appl
12	730.6	87.6	11076	4	US-09-539-601-25	Sequence 25, Appl
13	730.6	87.6	11076	4	US-09-539-601-31	Sequence 31, Appl
14	730.6	87.6	1037	1	US-08-462-195-1	Sequence 1, Appli
15	719.4	86.3	1037	2	US-08-636-883-1	Sequence 1, Appli
16	719.4	86.3	1037	1	US-09-127-829-1	Sequence 1, Appli
17	716.2	85.9	1037	2	US-08-324-977-9	Sequence 9, Appli
18	716.2	85.9	1167	2	US-08-384-616-9	Sequence 9, Appli
19	716.2	85.9	1167	3	US-08-904-686A-9	Sequence 9, Appli
20	716.2	85.9	1167	3	US-09-315-850-9	Sequence 9, Appli
21	716.2	85.9	1499	1	US-08-324-977-3	Sequence 3, Appli
22	716.2	85.9	1499	2	US-08-384-616-3	Sequence 3, Appli
23	716.2	85.9	1499	3	US-08-904-686A-3	Sequence 3, Appli
24	716.2	85.9	1499	3	US-09-315-850-3	Sequence 3, Appli
25	716.2	85.9	6039	1	US-08-324-977-11	Sequence 11, Appl
26	716.2	85.9	6039	2	US-08-384-616-11	Sequence 11, Appl
27	716.2	85.9	6039	2	US-08-904-686A-11	Sequence 11, Appl
28	716.2	85.9	6039	3	US-09-315-850-11	Sequence 11, Appli
29	716.2	85.9	9030	1	US-08-324-977-13	Sequence 13, Appl
30	716.2	85.9	9030	1	US-08-384-616-13	Sequence 13, Appl
31	716.2	85.9	9030	2	US-08-904-686A-13	Sequence 13, Appl
32	716.2	85.9	9030	2	US-09-315-850-13	Sequence 13, Appl
33	716.2	85.9	9416	1	US-08-324-977-1	Sequence 1, Appli
34	716.2	85.9	9416	2	US-08-384-616-1	Sequence 1, Appli
35	716.2	85.9	9416	3	US-08-904-686A-1	Sequence 1, Appli
36	716.2	85.9	9416	3	US-08-823-895A-27	Sequence 27, Appl
37	716.2	85.9	9416	1	US-08-462-195-3	Sequence 3, Appli
38	716.2	85.5	1037	1	US-08-636-883-3	Sequence 3, Appli
39	713	85.5	1037	3	US-09-127-829-3	Sequence 3, Appli
40	713	85.5	1037	3	US-08-150-204E-96	Sequence 96, Appl
41	711.4	85.3	9472	3	US-08-811-556-5	Sequence 5, Appli
42	594.6	71.3	12980	3	US-09-034-756-5	Sequence 5, Appli
43	594.6	71.3	12980	3	US-09-315-850-7	Sequence 7, Appli
44	593	71.1	9365	4	US-09-827-688-7	Sequence 7, Appli
45	593	71.1	9401	2	US-08-432-693-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-191-160-5
Sequence 5, Application US/08191160
Patent No. 6210675
; GENERAL INFORMATION:
; APPLICANT: Highfield, Peter Edmund
; APPLICANT: Rodgers, Brian Colin
; APPLICANT: Tedder, Richard Seton
; APPLICANT: Barbara, John Anthony James
; TITLE OF INVENTION: Viral Agent
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: 1700 K Street
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT compatible
; OPERATING SYSTEM: MS-DOS V3.2
; SOFTWARE: Wordperfect 5.0 (DOS text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/191,160
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/628,516
; FILING DATE: 17 DEC 1990
; APPLICATION NUMBER: UK 89 28 562.1
; FILING DATE: 18 DEC 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 90 04 414.0
; FILING DATE: 27 FEB 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 90 04 814.1
; FILING DATE: 03 MAR 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: E. Anthony Figg
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1645-103A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-5740
; TELEFAX: (202) 833-5744
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 834 base pairs
; TYPE: nucleotide with corresponding protein
; STRANDEDNESS: single

;
; MOLECULE TYPE: cDNA to genomic RNA
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: human; serum infectious for PT-NANBH
; IMMEDIATE SOURCE:
; LIBRARY: clone BR11 from cDNA library in lambda gt11
; FEATURE:
; LOCATION: from 1 to 834 bp portion of the PT-NANBH
; LOCATION: polyprotein
; OTHER INFORMATION: probably encodes viral structural
; OTHER INFORMATION: proteins
US-08-191-160-5

Query Match
Best Local Similarity   100.0%;   Score 834; DB 3; Length 834;
Matches 834; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGAGACGTCAGGTTCCCGGCGGTGGT    60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGAGACGTCAGGTTCCCGGCGGTGGT    60

Qy   61 CAGATCGTTGGTGGAGTTTACCTTGTTGCCGCGCGAGGGCCCCCAGGTTGGGTGTTGCGCCCG   120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAGATCGTTGGTGGAGTTTACCTTGTTGCCGCGCGAGGGCCCCCAGGTTGGGTGTTGCGCCCG   120

Qy  121 ACTAGGAAGAGCTTCCGGACGGGTCGCAAACCTCGTTGGAAGGGACAAACCTATTCCCCAAGGCT   180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACTAGGAAGAGCTTCCGGACGGGTCGCAAACCTCGTTGGAAGGGACAAACCTATTCCCCAAGGCT   180

Qy  181 CGCCAGCCCGAGGGCAGGGCCTGGGCGTACCCTTGGGCCCCTCTTATGGCAAC   240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CGCCAGCCCGAGGGCAGGGCCTGGGCGTACCCTTGGGCCCCTCTTATGGCAAC   240

Qy  241 GAGGGCATGGGGGTGGGCAGGATGGCTCCCTGTCACCCCGTGGCTCTCCGGCCTAGTTGGGGC   300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GAGGGCATGGGGGTGGGCAGGATGGCTCCCTGTCACCCCGTGGCTCTCCGGCCTAGTTGGGGC   300

Qy  301 CCCACTGACCCCCGGGCGTAGGTGCGGTAATTTGGGTAAAGTCATCGAATACCCTCACATGC   360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CCCACTGACCCCCGGGCGTAGGTGCGGTAATTTGGGTAAAGTCATCGAATACCCTCACATGC   360

Qy  361 GGCTTCGCCCGACTCTCATGGGGGTACATTCCGCTCGTCTCGGGGCGGCGCTGCC   420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GGCTTCGCCCGACTCTCATGGGGGTACATTCCGCTCGTCTCGGGGCGGCGCTGCC   420

Qy  421 AGGGCCCTGGCGCATGGGGTCCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAAT   480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGGGCCCTGGCGCATGGGGTCCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAAT   480

Qy  481 TTACCCGGTTGCTCTTTCTCTATCTTCTCTTGGCTTTGCTGGTCCTGTTGTTGACCATTCCA   540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TTACCCGGTTGCTCTTTCTCTATCTTCTCTTGGCTTTGCTGGTCCTGTTGTTGACCATTCCA   540

Qy  541 GCTTCCGCTTATGAAGTCGCAACGTGTCGGGGATCTACCATGTCACGGAACGGATTGCTCC   600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GCTTCCGCTTATGAAGTCGCAACGTGTCGGGGATCTACCATGTCACGGAACGGATTGCTCC   600

Qy  601 AACTCAAGCATCGTGTACGGAGACAGGCGGACAATCATCATGCACACACCCCGGGGTGTGCCC   660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 AACTCAAGCATCGTGTACGGAGACAGGCGGACAATCATCATGCACACACCCCGGGGTGTGCCC   660

Qy  661 TGTGTCCGGGAGGGTAATTCCTCCCGCGCTGCTGGGTAGCGCCGCTCACTCACTGCCGCC   720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TGTGTCCGGGAGGGTAATTCCTCCCGCGCTGCTGGGTAGCGCCGCTCACTCACTGCCGCC   720

Qy  721 AAGGACGGCCAGCATCCCACTCGGACAATACGACGCGCCACGTCGATTTGCTCGTTGGGCG   780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AAGGACGGCCAGCATCCCACTCGGACAATACGACGCGCCACGTCGATTTGCTCGTTGGGCG   780

Qy  781 GCTGCCCTTCTCGTCCGCTATGTACGTGGGGGGATCTCTCGGGGATCGTGTTTTCCCG   834
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GCTGCCCTTCTCGTCCGCTATGTACGTGGGGGGATCTCTCGGGGATCGTGTTTTCCCG   834
```

RESULT 2
US-08-191-160-21
; Sequence 21, Application US/08191160
; Patent No. 6216675
; GENERAL INFORMATION:
; APPLICANT: Highfield, Peter Edmund
; APPLICANT: Rodgers, Brian Colin
; APPLICANT: Tedder, Richard Seton
; APPLICANT: Barbara, John Anthony James
; TITLE OF INVENTION: Viral Agent
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: 1700 K Street
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT compatible
; OPERATING SYSTEM: MS-DOS V3.2
; SOFTWARE: Wordperfect 5.0 (DOS text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/191,160
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; APPLICATION NUMBER: 07/628,516
; FILING DATE: 17 DEC 1990
; APPLICATION NUMBER: UK 89 28 562.1
; FILING DATE: 18 DEC 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: .
; FILING DATE: 27 FEB 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 90 04 414.0
; FILING DATE: 27 FEB 1990
; APPLICATION NUMBER: UK 90 04 814.1
; FILING DATE: 03 MAR 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: E. Anthony Figg
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1645-103A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-5740
; TELEFAX: (202) 833-5744
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2116 base pairs
; TYPE: nucleotide with corresponding protein
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to genomic RNA
; ORIGINAL SOURCE:
; ORGANISM: human; serum infectious for PT-NANBH
; IMMEDIATE SOURCE:
; LIBRARY: cDNA clones from 5' end of the genome
; FEATURE:
; LOCATION: from 308 to 2116 bp start of the PT-NANBH
; LOCATION: polyprotein
; OTHER INFORMATION: viral structural and non-structural
; OTHER INFORMATION: proteins
US-08-191-160-21

Query Match
Best Local Similarity   96.2%;   Score 802.6; DB 3; Length 2116;
Matches 827; Conservative 0; Mismatches 4; Indels 0; Gaps 2;
US-08-191-160-21

```
Qy    1 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGAGACGTCAGGTTCCCGGCGGTGGT    60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  332 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGAGACGTCAGGTTCCCGGCGGTGGT   391

Qy   61 CAGATCGTTGGTGGAGTTTACCTTGTTGCCGCGCGAGGGCCCCCAGGTTGGGTGTTGCGCCCG   120
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         392 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 451

Qy         121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         452 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 511

Qy         181 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         512 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAC 571

Qy         241 GAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         572 GAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGC 631

Qy         301 CCCACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         632 CCCACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 691

Qy         361 GGCTTCGCCGACTCTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTA-GGGGCGCTGC 419
               |||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Db         692 GGCTTCGCCGAC-CTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTAGGGGGCGCTGC 750

Qy         420 CAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAA 479
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         751 CAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAA 810

Qy         480 TTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCC 539
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         811 TTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCC 870

Qy         540 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTC 599
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         871 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTC 930

Qy         600 CAACTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGTGTGTGCC 659
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         931 CAACTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGTGTGTGCC 990

Qy         660 CTGTGTCCGGGAGGGTAATTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGC 719
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         991 CTGTGTCCGGGAGGGTAATTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGC 1050

Qy         720 CAAGGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGGGC 779
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1051 CAAGGACGCCAGCATCCCCACTGCGACAATACGACGCCACGTCGATTTGCTCGTTGGGGC 1110

Qy         780 GGCTGCCTTCTCGTCCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCC 832
               |||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        1111 GGCTGCCTTCTGCTCCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCC 1163

RESULT 3
US-08-612-973-49
; Sequence 49, Application US/08612973
; Patent No. 6150134
;  GENERAL INFORMATION:
;    APPLICANT:  MAERTENS, GEERT
;    APPLICANT:  BOSMAN, FONS
;    APPLICANT:  DE MARTYNOFF, GUY
;    APPLICANT:  BUYSE, MARIE-ANGE
;    TITLE OF INVENTION:  PURIFIED HEPATITIS C VIRUS ENVELOPE
;    TITLE OF INVENTION:  PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,973
;      FILING DATE:  11-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BYRNE, THOMAS E.
;      REGISTRATION NUMBER:  32,205
;      REFERENCE/DOCKET NUMBER:  1487-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2433 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2430
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..2427
US-08-612-973-49

  Query Match             88.4%;  Score 737;  DB 3;  Length 2433;
  Best Local Similarity   94.4%;  Pred. No. 6.2e-215;
  Matches  786;  Conservative    0;  Mismatches   45;  Indels    2;  Gaps    2;

Qy           1 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCCGGGCGGTGGT 60
               ||||||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||
Db          25 AGAAAAACCAAACGTAACACCAACCGCCGCCCACAGGACGTCAAGTTCCCGGGCGGTGGT 84

Qy          61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          85 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 144

Qy         121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 180
               ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         145 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGGAGGCGACAACCTATCCCCAAGGCT 204

Qy         181 CGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAC 240
               ||||  |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         205 CGCCGACCCGAGGGTAGGGCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAAT 264

Qy         241 GAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGGCTCCCGGCCTAGTTGGGGC 300
               |||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||||
Db         265 GAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGGCTCTCGGCCTAGTTGGGGC 324

Qy         301 CCCACTGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360
               || || |||||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db         325 CCTACAGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTCATCGATACCCTTACATGC 384

Qy         361 GGCTTCGCCGACTCTCATGGGGTACATTCCGCTCGTCGGCGCTCCCTTA-GGGGCGCTGC 419
               |||||||||||| ||| ||||||||||||||||||||||||| ||| || ||||||||||
Db         385 GGCTTCGCCGAC-CTCGTGGGGTACATTCCGCTCGTCGGCGCCCCCCTAGGGGGCGCTGC 443

Qy         420 CAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAA 479
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         444 CAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAA 503

Qy         480 TTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTTTGACCATTCC 539
               ||| ||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||
Db         504 TTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCCTGTCTGACCGTTCC 563

Qy         540 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTC 599
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||||
```

Left column:

```
Db   564 AGCTTCCGCCTTATGAAGTGCGCAAACGTGTCCGGGGATGTGTACCATGTCACGGAACGACTGCTC
```

RESULT 4
US-08-927-597-49
; Sequence 49, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
;   APPLICANT: MAERTENS, GEERT
;   APPLICANT: BOSMAN, FONS
;   APPLICANT: DE MARTYNOFF, GUY
;   APPLICANT: BUYSE, MARIE-ANGE
;   TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
;   TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;   NUMBER OF SEQUENCES: 111
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: NIXON & VANDERHYE P.C.
;   STREET: 1100 NORTH GLEBE ROAD
;   CITY: ARLINGTON
;   STATE: VIRGINIA
;   COUNTRY: U.S.A.
;   ZIP: 22201-4714
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/927,597
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/612,973
;   FILING DATE: 11-MAR-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: BYRNE, THOMAS E.
;   REGISTRATION NUMBER: 32,205
;   REFERENCE/DOCKET NUMBER: 1487-10
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 816-4000
;   TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 49:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2433 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..2430
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: 1..2427
; US-08-927-597-49

Right column:

```
Qy   600 CAACTCAAGCATCGTGTACGAGACAGCGGGACATGATCATGTCGCACACCCCCCGGGTGTGTGCC 659
         ||||||||||  || ||| |||||||||||||||||||||||||  |||||||||||||||||||
Db   624 CAACTCAAGCATTGTGTATGAGAGCAGCGGGACATGATCATGCACACCCCCGGGGTGCGTGCC 683

Qy   660 CTGTGTCCGGGAGGGTAATTCCTCCCGCGTGCTGGGTAGCCGCTCACTCCCCACGGCTTCGCGGC 719
         ||||||||  |   | | |||  ||  | |||||||||| || |||||| |||| | ||||||||
Db   684 CTGCGTTCGGGAGAACAACTCTTCCCCGCTGCTGGGTAGCCGCTCACCCCCACCGCCTTCGCAGC 743

Qy   720 CAAGGACCGCCAGCGATCCCCACTCGCGACAATACGACGCCACGTCGCTCGATTTGCTCGTTGGGC 779
         ||   | |||| |  |||||||| |  |  ||||||||||||||||||||||||||||||||||||
Db   744 TAGGAACGGCCAGCGGTCCCCACCACGACAAATACGACGCCACGTCGCTCGATTTGCTCGTTGGGC 803

Qy   780 GGCTGCCTTTCTCGTCCGCTATGTACGGTGGGGGGGATCTTCTGCGGGATCTGTTTTTCC 832
         |  |  ||| ||||| | | ||  |  ||||||| |||||||||||||||||||||||||
Db   804 GGCTGCTTTTCTGTTCCGCTATGTACGTGGGGGGGACCTCTGCGGGATCTGTTCGTCTTCC 856
```

Best Local Similarity   88.4%;   Score 737; DB 3; Length 2433;
Matches 786; Conserve   0; Pred No. 6.2e-215;
                                Mismatches   45; Indels   2; Gaps   2;

RESULT 5
US-09-827-688-6
; Sequence 6, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
;   APPLICANT: ORSON, FRANK
;   APPLICANT: KINSEY, BERMA
;   APPLICANT: BIOGAL, BALBIR
;   TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION D
;   TITLE OF INVENTION: AGENTS
;   FILE REFERENCE: P01949US1/10004014

; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.0
;
; SEQ ID NO 6
; LENGTH: 9413
; TYPE: DNA
; ORGANISM: HEPATITIS C

US-09-827-688-6

Query Match   88.0%;   Score 733.8;   DB 4;   Length 9413;
Best Local Similarity 94.1%;   Pred. No. 1.2e-213;
Matches 784;   Conservative 0;   Mismatches 47;   Indels 2;   Gaps 2;

Qy   1 AGAAAAACCAAAACGTAACACCAACCGCCCCCCCACAGGACGTCAGGTTCCCGGCGTTGGT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   354 AGAAAAACCAAAACGTAACACCAACCGCCCCCCCACAGGACGTTAAGTTCCCGGCGTTGGT 413

Qy   61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGGTGTGCGCGCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 CAGATCGTTGGTGGAGTTTACCCTGTTGCCGCGCAGGGGCCCCAGGTTGGGGTGTGCGCGCG 473

Qy   121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTCGGAAGGCGACAACCTATCCCCAAGGCT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   474 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 533

Qy   181 CGCCAGCCCGAGGGCAGGGCCTCAGCCCGGGGTACCCCTTGGCCCCTCTATGGCAAC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   534 CGCCGGCCCGAGGGTAGGACCTGGGCTCAGCCCGGGGTACCCCTCTATGGCAAC 593

Qy   241 GAGGGCCATGGGGTGGGCAGGAGGCTCCTGTCACCCGTGGCTCCCGGCCTAGTTGGGGC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 GAGGGTATGGGGTGGGGCAGGATGGCTCCTCTGTCACCCGTGGCTCCCGGCCTAGTTGGGGC 653

Qy   301 CCCACTGACCCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   654 CCCACAGACACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTCATCGATACCCTTACATGC 713

Qy   361 GGCTTCGCGCGACTCTCATTGGGGTACATTCCGCGTCGGCGGCTCCTTA-GGGGCGCTGC 419
         |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   714 GGCTTCGCGGAC-CTCATGGGGGTACATTCCGCGTTGTCGGCGGCTGC 772

Qy   420 CAGGGCCCTGGCCGCATGGCGGTCCGGGTTCTGGAGGACGGCGTGAACTATGCAACAGGGAA 479
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   773 CAGGGCCCTGGCCACATGGTGTCCGGGTTCGGAGGACGGCGTGAACTATGCAACAGGGAA 832

Qy   480 TTTACCCGGGTTGCTCTTTCTCTATCTTCCTCTCTCGGCTTTGCTGTTTGACCATTCC 539
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   833 TCTGCCCGGGTTGCTCTTTCTCTATCTTCCTCTCTCGGCTTTGCTGTTTTGACCATTCC 892

Qy   540 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCGGGATCTACCATGTCACGAACGATTGCTC 599
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   893 AGCTTCCGCTTACGAGGTGCGCAACGTGTCCGGGATATACCATGTCACGAACGACTGCTC 952

Qy   600 CAACTCAAGCATCGTGTACGAGACAGCGGACATGATCATCATGAACTGGGGCATATGTGTGCC 659
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   953 CAACTCAAGTATTGTGTATGAGGCGGACATGATCATCATGAACTGGGGCGTTGTCGGGC 1012

Qy   660 CTGTGTTCCGGGAGGGGGTTATTCCTCCCGCGTGGGTAGGCGCTCCACTCGCTGCGGCC 719
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1013 CTTCGTCGGGGGAGAGTAATTTCTTCCTCCCGCGTGGGTAGGCGGTCCACTCGCTGCGGCC 1072

Qy   720 CAAGGACGCCAGCAATCCCACTGGGACAATACGACGCCACGTCGATTTGCTCGTTGGGGC 779
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1073 CAGGAACAGCAGCATCCCACCAGGACAATACGGACGCCACGTCGATTTGCTCGTTGGGGC 1132

Qy   780 GGCTGGCCTTCGTCCGCGTATATGTACGTGGGGAGGGATCTTCTCGGGATCCTGTTTTCC 832
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1133 GGCTGCTCTCGTTCCGCGTATGTACGTTGGGGGGATCTCTCGGGATCCGTTTTTC 1185

RESULT 6
US-08-470-426B-17
; Sequence 17, Application US/08470426B
; Patent No. 5856458
; GENERAL INFORMATION:
; APPLICANT: Okamoto, Hiroaki
; APPLICANT: Nakamura, Tetsuo
; TITLE OF INVENTION: OLIGONUCLEOTIDE PRIMERS, AND THEIR
; TITLE OF INVENTION: APPLICATION FOR HIGH-FIDELITY DETECTION OF NON-A, NON-B
; TITLE OF INVENTION: HEPATITIS VIRUS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Beveridge, DeGrandi, Weilacher & Young,
;   ADDRESSEE: L.L.P.
;   STREET: 1850 M Street, N.W., Suite 800
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20036
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/470,426B
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 536
;   FILING DATE: 12-JUN-1990
; ATTORNEY/AGENT INFORMATION:
;   NAME: Weilacher, Robert G.
;   REGISTRATION NUMBER: 20,531
;   REFERENCE/DOCKET NUMBER: 06/59-47083.1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 659-2811
;   TELEFAX: (202) 659-1462
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1539 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)

US-08-470-426B-17

Query Match   87.6%;   Score 730.6;   DB 2;   Length 1539;
Best Local Similarity 93.9%;   Pred. No. 4.4e-213;
Matches 782;   Conservative 0;   Mismatches 49;   Indels 2;   Gaps 2;

Qy   1 AGAAAAACCAAAACGTAACACCAACCGCCCCCCCACAGGACGTCAGGTTCCCGGCGTTGGT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 AGAAAAACCAAAACGTAACACCAACCGCCCCCCCACAGGACGTTAAGTTCCCGGCGTTGGT 84

Qy   61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGGTGTGCGCGCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAGATCGTTGGTGGAGTTTACCCTGTTGCCGCGCAGGGGCCCCACAGGTTGGGGTGTGCGCGCG 144

Qy   121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTCGGAAGGCGACAACCTATCCCCAAGGCT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGATGGGCGACAACCTATCCCCAAGGCT 204

Qy   181 CGCCAGCCCGAGGGCAGGGCCTCAGCCCGGGGTACCCCTTGGCCCCTCTATGGCAAC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 CGCCGACCCCAGGGCAGGGCCTCAGCCCGGGGTACCCCTTGGCCCCTCTATGGCAAT 264

Qy   241 GAGGGCCATGGGGTGGGCAGGAGGCTCCTGTCACCCGTGGCTCCCGGCCTAGTTGGGGC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265 GAGGGCCTTGGGGTGGGCAGGATGGCTCCTGTCACCCGTGGCTCCCGGCCTAGTTGGGGC 324

Qy   301 CCCACTGACCCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   325 CCCACTGACCCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC 360

```
Db   325  CCCACGGACCCCCGGCGCCGGCGGTAGGTCGCGGTAACTTGGGTAAGGTCATCGATACCCTTACATGC  384
Qy   361  GGCTTCGCCGACTTCATGGGGGTTACATTCCGCTCGTCGTGGGCGTCCCTTA-GGGGGCGCTGC  419
Db   385  GGCTTCGCCGGA-TCTCATGGGGGTATATTCCCCTCGTCGCGGGCGCCCCCTTAGGGGGGCGCTGC  443
Qy   420  CAGGGCGCCTGGGGCCATGGGGTTCCGGGGTTCGGGGGACGGGGCGTGAACTATGCAACAGGGAA  479
Db   444  CAGGGGCCTTGGGCACACGGTGTCCGGGGGTTCTGGAGGACGGGGGGTGAACTATGCAACAGGGAA  503
Qy   480  TTTTACCCGGTTGGTCTTCTTTTCTCTATCTTCCTTGGCTTTGGCTTTGGCTTGTTTTGACCATTCC  539
Db   504  CTTGCCCCGGTTGCTCTTTCTCTTATCTTCCTGTGCTCCTCGTTTTGACCATTCC  563
Qy   540  AGCTTCCGCCTTATGAAGTGCGCAACGGGGTGTCCGGGGTACATTCCGCTCGCTCGCTCGTCGCTC  599
Db   564  AGCTTCCGCCTTATGAAGTGCGCAACGGGTGTCCGGGGATATACCCATGTCACGGAACGACTGCTC  623
Qy   600  CAACTCAAGCAATCGTGTACGAGACAGCGGGACAGGGGCAATGATCATGACACCCCGGGGGGTGTGCC  659
Db   624  CAACTCAAGCAATTGTGTTATGAGGCGGAGCGGGCAATGATCATGCATACTCCGGGGGGTGTGCC  683
Qy   660  CTGTGTCCGGGGAGGGGTAATTCCTCCCCGCTTGCTCGGGTACGCCTCACTCCCCACCGCTCGCCGGC  719
Db   684  CTGCGTTCGGGGAGGACAACAGCTCCCGTTTGCTCGGGTAGCCCTCACTCCCACCGGCTCGGGGC  743
Qy   720  CAAGGACGCGCCAGCCATCCCCACTGCGACAAATACGACGCGCGATTTTGTCTCGTTGGGGC  779
Db   744  CAGGAATGCCAGGGTCCCCACTACGACAAATACGACGCCACGTGGACTTGCTCGTTGGGGC  803
Qy   780  GGCTGCGCCTTCTCGTCCCGCTATGTACGTGGGGGGGGGATCTCTGCGCGGATCTGTTTTCC  832
Db   804  GGCTGCGTTTCTCGTCCCGCTATGTACGGTGGGGGGGGGGATCTCTGCGCGGATCTGTTTTCC  856
```

```
Query Match           87.6%;  Score 730.6;  DB 2;  Length 1863;
Best Local Similarity 93.9%;  Pred. No. 4.9e-213;
Matches 782;  Conservative  0;  Mismatches 49;  Indels 2;  Gaps 2;

Qy     1  AGAAAAACCAAAACGTAACACCAACCTCCGCCCACAGGAGGAGGTCAGGTTCCCGGCGCTGC  60
Db   349  AGAAAAACCAAAAACGTAACACCAACCGCCACAGGAGGACGTCAAGTTCCCGGCGGGGT  408
Qy    61  CAGATCGTTTGGTGGAGTTTAACTGTTTGCCGCGGGCAGGTTGGGGTGTGCGCCCG  120
Db   409  CAGATCGTTGGTGGAGTTTAACCTGTTGCCGCGGGCAGGGGCCCACAGGTTGGGGTGTGCGCGGG  468
Qy   121  ACTAGGAAGACTTCCGGACGGGGTCGCAACCTCGTGGGAAGGCGACAACCTTATCCCAAGGCT  180
Db   469  ACTAGGAAGACTTCCGACGGGGTCGCAACCTCGTGGGAGAGGCGACAACCTTATCCCCAAGGCT  528
Qy   181  CGCCAGCCCGAGGGGCCAGGCCCTGGCCCTCAGCCGCCCGGGGTACCCCTCTATGGCAAC  240
Db   529  CGCCGACCCGAGGGGCCAGGCCCTGGGCTACGCGGGGTACCCCTTCTATGGCAAT  588
Qy   241  GAGGGCCATGGGGGTGGGGTGGGGCAGGATGGCCTCCTGTCACCCCGTCAGCCCTAGTTGGGGC  300
Db   589  GAGGGGCTTGGGGTGGGGCAGGATGGCCTCCTGTCACCCCGGGCCTAGTTGGGGC  648
Qy   301  CCCACTGACCCCCGGCGTAGGTCGCGGTAACTTGGGGTAAAGTCATCGATACCCTTACATGC  360
Db   649  CCCACGGACCCCCGGCGCCGGCGGTAGGTCGCGGTAACTTGGGTAAGGTCATCGATACCCTTACATGC  708
Qy   361  GGCTTCGCCGACTTCATGGGGGTTACATTCCGCTCGTCGTGGGCGTCCCTTA-GGGGGCGCTGC  419
Db   709  GGCTTCGCCGGA-TCTCATGGGGGTATATTCCCCTCGTCGCGGGCGCCCCCTTAGGGGGGCGCTGC  767
Qy   420  CAGGGCGCCTGGGGCCATGGGGTTCCGGGGTTCGGGGGACGGGGCGTGAACTATGCAACAGGGAA  479
Db   768  CAGGGGCCTTGGGCACACGGTGTCCGGGGGTTCTGGAGGACGGGGGGTGAACTATGCAACAGGGAA  827
Qy   480  TTTTACCCGGTTGGTCTTCTTTTCTCTATCTTCCTTGGCTTTGGCTTTGGCTTGTTTTGACCATTCC  539
Db   828  CTTGCCCCGGTTGCTCTTTCTCTTATCTTCCTGTGCTCCTCGTTTTGACCATTCC  887
Qy   540  AGCTTCCGCCTTATGAAGTGCGCAACGGGGTGTCCGGGGTACATTCCGCTCGCTCGCTCGTCGCTC  599
Db   888  AGCTTCCGCCTTATGAAGTGCGCAACGGGTGTCCGGGGATATACCCATGTCACGGAACGACTGCTC  947
Qy   600  CAACTCAAGCAATCGTGTACGAGACAGCGGGACAGGGGCAATGATCATGACACCCCGGGGGGTGTGCC  659
Db   948  CAACTCAAGCAATTGTGTTATGAGGCGGAGCGGGCAATGATCATGCATACTCCGGGGGGTGTGCC  1007
Qy   660  CTGTGTCCGGGGAGGGGTAATTCCTCCCCGCTTGCTCGGGTACGCCTCACTCCCCACCGCTCGCCGGC  719
Db  1008  CTGCGTTCGGGGAGGACAACAGCTCCCGTTTGCTCGGGTAGCCCTCACTCCCACCGGCTCGGGGC  1067
Qy   720  CAAGGACGCGCCAGCCATCCCCACTGCGACAAATACGACGCGCGATTTTGTCTCGTTGGGGC  779
Db  1068  CAGGAATGCCAGGGTCCCCACTACGACAAATACGACGCCACGTGGACTTGCTCGTTGGGGC  1127
Qy   780  GGCTGCGCCTTCTCGTCCCGCTATGTACGTGGGGGGGGGATCTCTGCGCGGATCTGTTTTCC  832
Db  1128  GGCTGCGTTTCTCGTCCCGCTATGTACGGTGGGGGGGGGGATCTCTGCGCGGATCTGTTTTCC  1180
```

RESULT 8
US-09-014-416-4
; Sequence 4, Application US/09014416

Patent No. 6153421
GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 4
; LENGTH: 9595
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-4

Query Match   87.6%; Score 730.6; DB 3; Length 9595;
Best Local Similarity   93.9%; Pred. No. 1.1e-212;
Matches 782; Conservative   0; Mismatches 49; Indels 2; Gaps 2;

```
Qy      1 AGAAAAAACAAAACGTAACACCAACCTTCCGCCCCACAGGACGTCAGGTTCCCGGGCGGGT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 AGAAAAAACAAAACGTAACACCAACCGCCCCACAGGACGTCAAGTTCCCGGGCGGGTGGT 425

Qy     61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 485

Qy    121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    486 ACTAGGAAGGCTTCCGAGCGGTCGCAACCTATCCCAAAGGCT 545

Qy    181 CGCCAGCCCGAGGGGCCAGGGCCTCAGCCGGGCGTACCCTTGGCCCCTTCTATGGCAAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546 CGCCAGCCCGAGGGGCCAGGGCCTCAGCCGGGCGTACCCTTGGCCCCTTCTATGGCAAT 605

Qy    241 GAGGGCATGGGGTCGGGCAGGATGGCTCCTCGTGCACCCGGGTCTACCCTTGGCTCCGGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 GAGGGCCTTGGGGTGGGCAGGATGGCTCCTCGTGCACCCGGGTCTACCCTTGGCTCCGGG 665

Qy    301 CCCACTGACCCCGGCGTAGGTCGGCGTAATTTGGGTAAAGTCATCGGATACCCTCACATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 CCCACGGACCCCGGCGTAGGTCGGCGTAACCTTGGGTAAGTCATCGGATACCCTCACATGC 725

Qy    361 GGCTTCGGCCGACTTCTCATGGGGGTACATTCCGCGTCGGTAAAGTCATCATCGGATACGCC 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    726 GGCTTCGCCGA-TCTCATGGGGGTACATTCCGCGTCGGTAAGTCATCGGATACCCTACATGC 784

Qy    420 CAGGGCCCTGGCGCATGGCGTCCGGTGAACGTGAACACAGAGGGA 479
          |||||||||||||||||||||||||||||||||||||||||||||
Db    785 CAGGGCCTTGGCACACGGTGTCCGGGTTCGGAGGACGGGCTGAACTATGCAACAGGGAA 844

Qy    480 TTTACCCGGTTGCTCTTTCTCTCATTCTTCTTGGCTTTGCTCGTGTTTGACCATTCC 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    845 CTTGCCCGGTTGCTCTTTCTCTATCTTTCTCTTGGCTTTGCTCGTGTTTGACCATTCC 904

Qy    540 AGCTTCCGCTTATGAAGTCGCAGCAGCACATGACACACCCACCCGGGGTGTGTGCC 599
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    905 AGCTTCCGCTTATGAAGTCGCAGCAGCACATGACACACCCACCCGGGGTGTGTGCC 964

Qy    600 CAACTCAACCATCGTGTACGAGACAGCGCATGACACCACCCGGGGTGTGTGCC 659
          |||||||||||||||||||||||||||||||||||||||||||||
Db    965 CAACTCAAGCATTGTATGAAGGAGCCAGCGGACGGACTGATCATGCATAGTGCGGGGTTGCC 1024

Qy    660 CTGTGTCGGGGAGGGGTAATTCCTCCGCGTTGCTGCTGGTGGGGTAACAGAGCTCCCGTTGCTGGGGCTGCGGC 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1025 CTGTGTTCAGGGAGGGTAACAGAGCTCCGGTTGCTCACTCCACGCGTTGCTGGGGC 1084
```

RESULT 9
US-09-014-416-6
Sequence 6, Application US/09014416
Patent No. 6153421
GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 6
; LENGTH: 9599
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-6

Query Match   87.6%; Score 730.6; DB 3; Length 9599;
Best Local Similarity   93.9%; Pred. No. 1.1e-212;
Matches 782; Conservative   0; Mismatches 49; Indels 2; Gaps 2;

```
Qy      1 AGAAAAAACAAAACGTAACACCAACCTTCCGCCCCACAGGACGTCAGGTTCCCGGGCGGGT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 AGAAAAAACAAAACGTAACACCAACCGCCCCACAGGACGTCAAGTTCCCGGGCGGGTGGT 425

Qy     61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 485

Qy    121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||
Db    486 ACTAGGAAGGCTTCCGAGCGGTCGCAACCTATCCCAAAGGCT 545

Qy    181 CGCCAGCCCGAGGGGCCAGGGCCTCAGCCGGGCGTACCCTTGGCCCCTTCTATGGCAAC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546 CGCCAGCCCGAGGGGCCAGGGCCTCAGCCGGGCGTACCCTTGGCCCCTTCTATGGCAAT 605

Qy    241 GAGGGCATGGGGTCGGGCAGGATGGCTCCTCGTGCACCCGGGTCTACCCTTGGCTCCGGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 GAGGGCCTTGGGGTGGGCAGGATGGCTCCTCGTGCACCCGGGTCTACCCTTGGCTCCGGG 665

Qy    301 CCCACTGACCCCGGCGTAGGTCGGCGTAATTTGGGTAAAGTCATCGGATACCCTCACATGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 CCCACGGACCCCGGCGTAGGTCGGCGTAACCTTGGGTAAGTCATCGGATACCCTCACATGC 725

Qy    361 GGCTTCGGCCGACTTCTCATGGGGGTACATTCCGCGTCGGTAAAGTCATCATCGGATACGCC 419
          ||||||||||||||||||||||||||||||||||||||||||||
Db    726 GGCTTCGCCGA-TCTCATGGGGGTACATTCCGCGTCGGTAAGTCATCGGATACCCTACATGC 784

Qy    420 CAGGGCCCTGGCGCATGGCGTCCGGTGAACGTGAACACAGAGGGA 479
          |||||||||||||||||||||||||||||||||||||||||
Db    785 CAGGGCCTTGGCACACGGTGTCCGGGTTCGGAGGACGGGCTGAACTATGCAACAGGGAA 844

Qy    480 TTTACCCGGTTGCTCTTTCTCTCATTCTTCTTGGCTTTGCTCGTGTTTGACCATTCC 539
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    845 CTTGCCCGGTTGCTCTTTCTCTATCTTTCTCTTGGCTTTGCTCGTGTTTGACCATTCC 904
```

Qy   720 CAAGGACGCGCAGCAATCCCCACTGCGACAATACGACGCCCAGCGTTGGGGC 779
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1085 CAGGGAATGCCAGCGTCCCCCACTACCACCAATACCGACGCCCAGGAC 1144

Qy   780 GGCTGCGCCTTCTCGTCCGGCTATGTACGGTGGGGGATCTCTGGGGGATCTGTTTTCC 832
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1145 GGCTGCTTTTCTGCCGGTATGTACGGTGGGGGATCTTCTGGGGACTCTATTTTCC 1197

RESULT 10
US-09-539-601-1

; Sequence 1, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 11076
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/Core-3'/wt
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION: internal ribosome entry site from
; OTHER INFORMATION: encephalomyocarditis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(10845)
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: nonstructural protein NS5B; parental sequence
; OTHER INFORMATION: without cell culture-adaptive mutations
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)

Query Match   87.6%; Score 730.6; DB 4; Length 11076;
Best Local Similarity   93.9%;
Matches 782; Conservative   0; Mismatches 49; Indels 2; Gaps 2;

Qy   1 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCGGGCGGGTGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1837 AGAAAAACCAAACGTAACACCAACCGCGCCCCAGGACGTCAAGTTCCCGGGCGGGTGGT 1896

Qy   61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCGCGCG 120

; OTHER INFORMATION: construct I389/Core-3'/9-13F
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION:
; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: nonstructural protein NS5B; carries cell
; OTHER INFORMATION: culture-adaptive mutations from clone 9-13F
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)
US-09-539-601-19   LOCATION: (1813)..(10845)

Query Match   87.6%;   Score 730.6;   DB 4;   Length 11076;
Best Local Similarity   93.9%;   Pred. No. 1.2e-212;
Matches 782;   Conservative   0;   Mismatches   49;   Indels   2;   Gaps   2;

```
Qy     1 AGAAAAACCAAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCGGGCGGT   60
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  1837 AGAAAAACCAAAACGTAACACCAACCGCCCCACAGGACGTCAAGTTCCGGGCGGTGGT 1896

Qy,   61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCACTACGACCTCAAGTTTGGGTGTGCCGCGCGG  120
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1897 CAGATCGTCGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCCGCGCGG 1956

Qy   121 ACTAGGAAGACTTCCGAGCGGGTCCCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1957 ACTAGGAGAGACTTCCGAGCGGGTCCCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 2016

Qy   181 CGCCAGCCCGAGGGCAGGGGCCCTGGGCGTACCCTTGGCCCCCTTCTATGGCAAC  240
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  2017 CGCCAGCCCGAGGGGTAGGCGCTCAGCCCGGGTACCCCTTCTATGGCAAT 2076

Qy   241 GAGGGGCATGGGGTGGGGCAGGATGGCTCTCGTCTGCTCCCGGCTCCCGGCTAGTTGGGGC  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2077 GAGGGCTTGGGGTGGGGCAGGATGGCTCTCCGTCTCACCCGGCTCCCGGCTAGTTGGGGC 2136

Qy   301 CCCACTGACCCCCGGCGTAGGTCCCGTAATTGGGTAAAGTCATCATGGGTCATCGATACCCTCACATGC  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2137 CCCACGGGACCCCCGGCGTAGGTCCCGCAATTTCGGTAAGGTCATCGATACCCTCACGTAC 2196

Qy   361 GGCTTCGCCGACTCTCATGGGTACATTCCGCTCGTCGCGGCTCCCTTA-GGGGCGGCTGC  419
           ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  2197 GGCTTCGCCGGA-TCTCATGGGGTACATTCCGCTCGTCGCGGCTCCCTAGGGGGCGGCTGC 2255

Qy   420 CAGGGCCCTGGGCCCATGCGGTCCCGGGTTCTGGAGGACCGGGTGAAACTACGCAACAGGGAA  479
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2256 CAGGGCCCCTGGGCGCCGTCCGGGGTTCTGGAGGACGGCGTGAACTATCCAACAGGGAA 2315

Qy   480 TTTACCCGGTTGTCTTTTCTCTATCTTCCTTGGCTTTTGCTCTGTTGTTTTGACCATTCC  539
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2316 TCTTGCCCGGTTGTCTTTTCTATCTTCTTTTTGGCTTTTGCTCTGTTGTTTGACCATTCC 2375

Qy   540 AGCTTCCGCTTATGAAGTCGGCGAACAGTGTCCGGGATCTACCATGTCACGAACGATTGCTC  599
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2376 AGCTTCCGCTTATGAAGTCGGCGAACAGTATCCGGGACTGTACCACATGTCACGAACGACTGCTC 2435

Qy   600 CAACTCAAGCCATCGTGTACGAGACAGCGGACAACATATCAATCACTGCACACCCCCGGGGTGTGTGCC  659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2436 CAACGCAAGCCATTGTGTATGAGGCAGACAGCATTGTCATCCCGGGTGCGTGCC 2495

Qy   660 CTGTGTCTCCGGGAGGGTAATTCCTCCCGCGTTGCCTGCCTCACTCCCACGCGTTCGCGCGGC  719
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2496 CTTCGGTTCGGGGAGAACACTCCTCCCGCGTGCTACTCCCACGCGTTCGCGCGGC 2555
```

; OTHER INFORMATION: construct I389/Core-3'/5.1
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/Core-3'/5.1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION:
; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(10845)
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: nonstructural protein NS5B; carries cell
; OTHER INFORMATION: culture-adaptive mutations of clone 5.1
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)
US-09-539-601-25   LOCATION: (10846)..(11076)

RESULT 12
US-09-539-601-25
; Sequence 25, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
; LENGTH: 11076
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)

Query Match   87.6%;   Score 730.6;   DB 4;   Length 11076;
Best Local Similarity   93.9%;   Pred. No. 1.2e-212;
Matches 782;   Conservative   0;   Mismatches   49;   Indels   2;   Gaps   2;

```
Qy     1 AGAAAAACCAAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCGGGCGGT   60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  1837 AGAAAAACCAAAACGTAACACCAACCGCCCCACAGGACGTCAAGTTCCGGGCGGTGGT 1896

Qy    61 CAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCACTACGACCTCAAGTTTGGGTGTGCCGCGCGG  120
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  1897 CAGATCGTCGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTGGGTGTGCCGCGCGG 1956

Qy   121 ACTAGGAAGACTTCCGAGCGGGTCGCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1957 ACTAGGAGAGACTTCCGAGCGGGTCCCAACCTCGTGGAAGGCGACAACCTATCCCCAAGGCT 2016

Qy   181 CGCCAGCCCGAGGGCAGGGGCCCTGGGCGTACCCTTGGCCCCCTTCTATGGCAAC  240
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  2017 CGCCAGCCCGAGGGGTAGGCGCTCAGCCCGGGTACCCTTGGCCCCCTTCTATGGCAAT 2076
```

RESULT 13

US-09-539-601-31

Sequence 31, Application US/09539601C

Patent No. 663343

GENERAL INFORMATION:

APPLICANT: Bartenschlager, Ralf FW

TITLE OF INVENTION: Hepatitis C Virus Cell Culture System

FILE REFERENCE: all sequences

CURRENT APPLICATION NUMBER: US/09/539,601C

CURRENT FILING DATE: 2001-08-30

EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY

EARLIER FILING DATE: 1999-04-03

NUMBER OF SEQ ID NOS: 51

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 31

LENGTH: 11076

TYPE: DNA

ORGANISM: Hepatitis C virus

FEATURE:

NAME/KEY: 5'UTR

LOCATION: (1)..(341)

OTHER INFORMATION:

FEATURE:

NAME/KEY: CDS

LOCATION: (342)..(1193)

OTHER INFORMATION: hepatitis C virus core - neomycin

OTHER INFORMATION: phosphotransferase fusion protein

FEATURE:

NAME/KEY: RBS

LOCATION: (1202)..(1812)

OTHER INFORMATION: internal ribosome entry site from

OTHER INFORMATION: encephalomyocarditis virus

FEATURE:

NAME/KEY: CDS

LOCATION: (1813)..(10845)

OTHER INFORMATION: hepatitis C virus polyprotein from core to

OTHER INFORMATION: nonstructural protein NS5B; carries cell culture

OTHER INFORMATION: adaptive mutations from clone no. 19

FEATURE:

NAME/KEY: 3'UTR

LOCATION: (10846)..(11076)

US-09-539-601-31

Query Match   87.6%;   Score 730.6;   DB 4;   Length 11076;
Best Local Similarity   93.9%;   Pred. No. 1.2e-212;
Matches   782;   Conservative   0;   Mismatches   49;   Indels   2;   Gaps   2;

[Sequence alignment data — query (Qy) and database (Db) nucleotide sequences with position numbers]

RESULT 14

US-08-462-195-1
Sequence 1, Application US/0846219S
Patent No. 5789544
  GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: MATSUURA, YOSHIHARU
; APPLICANT: HONDA, YOSHIKAZU
; APPLICANT: SEKI, MAKOTO
; TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF
; TITLE OF INVENTION: HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,195
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,303
; FILING DATE: 22-MAY-1995
; APPLICATION NUMBER: US 08/074,584
; FILING DATE: 11-JUN-1993
; APPLICATION NUMBER: JP 152487/1992
; FILING DATE: 11-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, P.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4169-003-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1037 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis C virus
; IMMEDIATE SOURCE:
; CLONE: pUC010
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 17..1036

US-08-462-195-1

Query Match   86.3%; Score 719.4; DB 1; Length 1037;
Best Local Similarity   93.0%; Pred. No. 9.7e-210;
Matches 775; Conservative   0; Mismatches 56; Indels 2; Gaps 2;

Qy   1 AGAAAAACCAAAACGTAACACCAACCTCCGCCGCCACAGGACGTTCAGGTTCCCGGGGCGGT   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGAAAAACCAAAACGTAACACCAACCTCCGCCGCCACAGGACGTTCAGGTTCCCGGGGCGGT   60

Qy   61 CAGATCGTTGGTTGGAGTTTACCTGTTGCCGCGCGCAGGGGGCCCCAGGTTGGGTGTCCGCGCG   120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 AGAAAAACCAAAACGTAACACCAACCTCCGCCCACAGGACGTTAAGTCCCGGGTGTGGT   100

Qy   101 CAGATCGTCGTTGGAGTTTACTTGTTGCCGCGCAGGGGGCCCCAGGTTGGGTGTGCGCGCG   160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 CAGATCGTCGTTGGAGTTTACTTGTTGCCGCGCAGGGGGCCCCAGGTTGGGTGTGCGCGCG   160

Qy   121 ACTAGGAAGACTTCCGAGCGGTCGCAAACCTCGTGTGGAAGGGGCGACAACCTATCCCAAGGCT   180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 ACTAGGAAGAACTTCCGAGCGGTCGCAACCTCGTGGAAGGGCGACAACCTATCCCAAGGCT   220

Qy   181 CGCCAGCCCCGAGGGCAGGAGGCCTGGGCGTCAGCCGCGGTTGGCCCCTCTATGGCCAAC   240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 CGCCGCCCCGAGGGGCAGGAACCTGGGCGTCAGCCGTATCCTTGGCCCCCTCTATGGCCAAT   280

Qy   241 GAGGGCATGGGGGTGGGGGGCGAGGAAGGGTCCTGTCACACCCCGTGGGCTCCCGGGCTGGGGC   300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 GAGGGCTTGGGGTGGGGGGCCAGGATGGGCTTCCTGTCACCCCGTGGGGCTCCCGGGCTGGGGC   340

Qy   301 CCCACTGACACCCCCGCGGGTTAGGTTCGGTCGTTAATTTGGGTAAAGTCATCACATACCCTCACATGC   360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   341 CCTAATGACACCCCCGGCGGTTAGGTTCGGTCGGTTAATTTGGGTAAAGTCATCACCCTTACATGC   400

Qy   361 GGCTTGCGCCGACTCTCATGGGGGTACATTCCCGGTCTCCCTTA-GGGGCGGCCTGC   419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   401 GGCTTGGCCCGAC-CTCATGGGGGTACATCCCGGCTTCTGCGGGGGCGGGGGCTGC   459

Qy   420 CAGGGGCCCTGGCGCATGGCCGTCCCGGGGTTCTGGAGGACGGGGGTGAACTATGCAAACAGGGAA   479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   460 CAGGGGCCCTGGCCACATGGCCGTGTCCCGGGGTTCGGGAGGACGGGGTGAACTATGCAAACAGGGAA   519

Qy   480 TTTACCCGGTTGCTCTTCTTCTCTATCTTCCCTTGCTTTGCTTGTTTGCTTGTTTGACCATTCC   539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   520 TTTGCCCGGTTGCTCTTTCTCTATCTTCCCTTAGCTCGTCGCTTGTTTGACCATCCC   579

Qy   540 AGCTTCCGCTTATGAAGTGGCCAACGTGTCCGGGGATCTACCAGTGTCACGGAACGAATGCTC   599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   580 AGCTTCCGCTTATGGAAGTGGCCAAGTGTCCGGGGATATACCATGTCACACGAACGTGCTC   639

Qy   600 CAACTCAAGCATCGTGTACGAGACAGCGGACATGATCATGCACACCCCCGGGGTGTGTGCC   659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   640 CAACTCAAGCATTGTGTATGAGGCCGGGAGGGTGATCATGCACGCCCCCGGGGTGGTGGCC   699

Qy   660 CTGTGTCCGGGAGGGTGTAATTCCTCCCGCGTCGCTCGTTGGGTAGCCGTCTCACTCGCTGCGGC   719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   700 CTGCGTTTCGGGAGGAACAAATTCCTCCCGGTTGCTCGGCAGCCGTCCACTCCCACGGCTCGCGGC   759

Qy   720 CAAGGACGACGCCAGCATCCCACTGCGACAAATACGACGCCACGTCGATTGCTTGGGGGC   779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   760 CAGGAATGCCAGCCGTCCCACTACGACGCCACAGTCGACTTGCTCGTTGGGGAC   819

Qy   780 GGCTGCGCTTCTCGTCGCTATGTAGCTGTGGGGGACTCTCTGCCGGATCTGTTTTTCC   832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   820 GGCTGCTTCTGCCTCCGCTATGTACGTGGGGGGGATCTCTGCCGGATCTGTTTTTCC   872

RESULT 15
US-08-636-883-1
; Sequence 1, Application US/08636883
; Patent No. 5830691
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: MATSUURA, YOSHIHARU
; APPLICANT: HONDA, YOSHIKAZU
; APPLICANT: SEKI, MAKOTO
; TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF
; TITLE OF INVENTION: HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```
;
; OPERATING SYSTEM:    PC-DOS/MS-DOS
; SOFTWARE:            Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/636,883
;   FILING DATE:
;   CLASSIFICATION:      435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/446,303
;   FILING DATE:
;   APPLICATION NUMBER:  US 08/074,584
;   FILING DATE:         11-JUN-1993
;   APPLICATION NUMBER:  JP 152487/1992
;   FILING DATE:         11-JUN-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME:                Oblon, No. 5830691man P.
;   REGISTRATION NUMBER: 24,618
;   REFERENCE/DOCKET NUMBER:  4169-003-0
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (703) 413-3000
;   TELEFAX:    (703) 413-2220
;   TELEX:      248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1037 base pairs
;     TYPE:    nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   ANTI-SENSE:  NO
;   ORIGINAL SOURCE:
;     ORGANISM:  Hepatitis C virus
;   IMMEDIATE SOURCE:
;     CLONE:  pUC010
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  17..1036
US-08-636-883-1
```

```
Query Match    86.3%;  Score 719.4; DB 2;  Length 1037;
Best Local Similarity  93.0%;  Pred. No. 9.7e-210;
Matches  775;   Conservative  0;  Mismatches  56;  Indels  2;  Gaps  2;

Qy      1 AGAAAAACCAAACGTAACACCAACCTCCGCCCACAGGACGTCAGGTTCCCGGCGTGGT    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AGAAAAACCAAACGTAACACCAACCGTCGCCCACCACAGGACGTTAAGTTCCCGGCGTGGT   100

Qy     61 CAGATCGTTGGTGGAGTTTACCTGTTGCGCGCCGCAGGGGCCCCAGGGTTGGGTGTTGCGCGCG   120
          |||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db    101 CAGATCGTTCGGTGGAGTTTACTTGTTGCCGCGCCGCAGGGGCCCCAGGGTTGGGTGTGCGTCCG   160

Qy    121 ACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTTGGAAGGGACAACCTATCCCCAGGCT   180
          |||||||||||||||||||||||||||||||||||||||||||
Db    161 ACTAGGAAGACTTCCGACGGGTCGCCACCTCGTTGGAAGGCGACAACCTATCCCCAGGCT   220

Qy    181 CGCCAGCCCCGAGGGCCAGGGCAGGAGGGCTCGGGCTCAGGCCCGGGGTACCTTCTATGGCAAC   240
          |||||||||||||||||||||||||||||||||||||||||||
Db    221 CGCCGGCCCCGAGGGCCAGGACCTGGGCCTCAGCCGTATCCTTCTATGGCAAT   280

Qy    241 GAGGGCCATGGGGGTCGGGCAGGATGGCTCTCGTTCATCCCCCGGCTTCTCCCGGCCTAGTTGGGGGC   300
          |||||||||||||||||||||||||||||||||||||||||||
Db    281 GAGGGCTTGGGGGTGGGGCAGGGATGGCTGGTAGGTTCGCGGTTCACCCCCGCGGCTTCCCGGCGTTGGGGC   340

Qy    301 CCCACTGACCCCCCGGCGT.GGTCGCGTAATTTGGGTAAAGTCATCGATACCCTCACATGC   360
          |||||||||||||||||||||||||||||||||||
Db    341 CCTAATGACCCCCGGCGGTAGGTCGGGTAGGTTCATCGATACCCTCACATGC   400

Qy    361 GGCTTCGCGCGACTCTCATGGGGGGTACATTCCGTCCGTCGTCGTCCCTTA-GGGGCGGCTGC   419
          |||||||||||||||||||||||||||||||||||||||||||
Db    401 GGCTTCGCCGAC--CTCATGGGGGTACATCCCCGGTTCTGGGAGGCGCGTGAACTATGCAACACAGGGAA   459

Qy    420 CAGGGCCCCTGGCACCATGGTGTCCGGGTTCTTGGGAGGACCGGGTCGAACTATGCAACACAGGGAA   479
          |||||||||||||||||||||||||||||||||||||||||||
Db    460 CAGGGCCCCTGGCACATGGTGTCCGGGTTCTTGGGAGGACCGGGTCGAACTATGCAACACAGGGAA   519
```

```
Qy    480 TTTACCCCGGTTGCTCTTTTTCTCTATCTTCCTCTTGGCTTTGCTTGTCCCTGTTTGTTGACCATTCC   539
          |||||||||||||||||||||||||||||||||||||||||||
Db    520 TTTGCCCCGGTTGCTCTTTTTCTCTATCTTCCTCTTAGCTCTTGCTGTCCCTGTTTGTTGACCATCCC   579

Qy    540 AGCTTCCCGCTTATGAAGTGCGCAACGTGTCCCGCGGGGATTCTACCATGTCACGGAACGATTTGCTC   599
          |||||||||||||||||||||||||||||||||||||||||||
Db    580 AGCTTCCGCTTATGAAGTGCGCAACGTGTCCCGGGGATAATACCATGTCACAAAACGATCGCTC   639

Qy    600 CAACTCAAGCATCGTGTACGGAGCAGCGGGACATGATCATGCACACCCCGGGGTTGGGTGTGCC   659
          |||||||||||||||||||||||||||||||||||||||||||
Db    640 CAACTCAAGCATTGTGTRATGAGGCGGGGCGGGGACGTGATTCATGCACACCCCGGGGTTGGGTGTGCC   699

Qy    660 CTGTGTCCGGGAGGGGTAARTTCCTCCCGCTGCTGGGTAGGCGTCACTCCCACCGCTCGCGCGC   719
          |||||||||||||||||||||||||||||||||||||||||||
Db    700 CTGCGGTTCGGGGAGGAACAATTCCTCCCGGTTGCTGGGGTAGGGCTCACTCCCCACGCTCGCGCGC   759

Qy    720 CAAGGACGCCCAGCAATCCCCACTGCGACAAATAACGACGTCCCACGTGCGATTTGCTCGTTCGTTCGGGC   779
          |||||||||||||||||||||||||||||||||||||||||||
Db    760 CAGGAAATGCCAGCGTCCCCACTACGACAATTACGACGCCACGTGCAACTTGCTTCGGTTGGGGAC   819

Qy    780 GGCTGCCTTCTCCGTCCGCTTAGTGTACGGTGGGGGGGATCCTGGCGGATCTGTTTTCC   832
          ||||  |||||||||||||||||||||||||||||||||||||||
Db    820 GGCTGCTTTTCTTGCTCCGGCTATGTAACGTGGGGGGGGATCTCTGCGGGGATCTGTTTTTCC   872
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 04:09:06 ; Search time 5144 Seconds
                                       (without alignments)
                                       10427.668 Million cell updates/sec

Title: US-09-664-363-19
Perfect score: 1107
Sequence: 1 TCCTCCCGCTGCTGGGTAGC.........GCGGGTTCGGGGGGCCTTGGTTG 1107
Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0
Searched: 4708233 seqs, 2422760795 residues
Total number of hits satisfying chosen parameters: 9416466
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Database :   GenEmbl:*
         1:   gb_ba:*
         2:   gb_htg:*
         3:   gb_in:*
         4:   gb_om:*
         5:   gb_ov:*
         6:   gb_pat:*
         7:   gb_ph:*
         8:   gb_pl:*
         9:   gb_pr:*
        10:   gb_ro:*
        11:   gb_sts:*
        12:   gb_sy:*
        13:   gb_un:*
        14:   gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1107	100.0	1107	6	A28151	A28151 PT-NANBH mR
2	1107	100.0	1107	6	A32200	A32200 NANBHV PT p
3	1107	100.0	1107	6	AR144048	AR144048 Sequence
4	1107	100.0	2116	6	A28155	A28155 PT-NANBH mR
5	1107	100.0	2116	6	A32202	A32202 NANBHV PT p
6	1107	100.0	2116	6	AR144050	AR144050- Sequence
7	950.2	85.8	9410	14	D50481	D50481 Hepatitis C
8	946	85.5	9359	14	AF313916	AF313916 Hepatitis
9	940.6	85.0	9410	14	HPCK1R2	HPCK1R2 Hepatitis
10	934.2	85.0	9410	14	HPCK1S2	HPCK1S2 Hepatitis
11	934.2	84.4	1280	14	D50485	D50485 Hepatitis C
12	934.2	84.4	1280	14	AY746700	AY746700 Hepatitis
13	934.2	84.4	1734	12	AY746693	AY746693 Hepatitis
14	934.2	84.4	1734	12	AY545952	AY545952 Synthetic
15	933.2	84.3	1734	12	AY545953	AY545953 Synthetic
16	927.8	83.8	1280	14	AY545951	AY545951 Synthetic
17	927.8	83.8	1280	14	AY746691	AY746691 Hepatitis
18	927.8	83.8	1280	14	AY746694	AY746694 Hepatitis
19	927.8	83.8	9033	14	HCJ238800	HCJ238800 Hepatitis
20	927.8	83.8	9585	14	AB049095	AB049095 Hepatitis
21	927.8	83.8	9605	14	AX739971	AX739971 Sequence
22	927.8	83.8	9605	14	AJ238799	AJ238799 Hepatitis
23	927.8	83.8	11076	6	AR406040	AR406040 Sequence
24	927.8	83.8	11076	6	AR406046	AR406046 Sequence
25	927.8	83.8	11076	6	AR406048	AR406048 Sequence
26	927.8	83.8	11076	6	AR406050	AR406050 Sequence
27	927.8	83.8	11076	6	AX036258	AX036258 Sequence
28	927.8	83.8	11076	6	AX036252	AX036252 Sequence
29	927.8	83.8	11076	6	AX036260	AX036260 Sequence
30	927.8	83.8	11076	6	AX036262	AX036262 Sequence
31	926.2	83.7	9379	14	AF207760	AF207760 Hepatitis
32	926.2	83.7	9431	14	HCU45476	HCU45476 Hepatitis C
33	924.6	83.5	9377	14	AF207756	AF207756 Hepatitis
34	924.6	83.5	9379	14	AF207766	AF207766 Hepatitis
35	923.6	83.4	9547	14	AB049091	AB049091 Hepatitis
36	923.6	83.4	9410	14	HPCK1R1	HPCK1R1 Hepatitis
37	921.4	83.2	9379	14	AF207761	AF207761 Hepatitis
38	919.4	83.1	2540	6	D50480	D50480 Hepatitis C
39	919.4	83.1	2540	6	E04260	E04260 cDNA encodi
40	919.4	83.1	2540	6	E07391	E07391 cDNA encodi
41	919	83.0	9448	14	E04805	E04805 cDNA to 5'-
42	919	83.0	9580	14	HPC0483	HPC0483 Hepatitis C
43	918.2	82.9	9580	14	AF054250	AF054250 Hepatitis
44	918.2	82.9	1280	14	AY746696	AY746696 Hepatitis
45	918.2	82.9	9456	14	D10334	D10334 Hepatitis C
	918.2	82.9	9598	14	AB049101	AB049101 Hepatitis

ALIGNMENTS

RESULT 1
A28151

```
LOCUS       A28151                   1107 bp    DNA      linear   PAT 07-JUN-1995
DEFINITION  PT-NANBH mRNA fragment from patent GB2239245.
ACCESSION   A28151
VERSION     A28151.1  GI:1248634
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1107)
  AUTHORS
  TITLE
  JOURNAL   Patent: GB 2239245-A 19 26-JUN-1991;
            Post-transfusional non-A non-B hepatitis viral polypeptides
FEATURES             Location/Qualifiers
     source          1..1107
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /clone="136/155"
     CDS             <1..>1107
                     /codon_start=1
                     /transl_table=11
                     /product="PT-NANBH polyprotein"
                     /protein_id="CAA01929.1"
                     /db_xref="GI:1248635"
                     /translation="SSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMYVG
                     DLCGSVFLVSQLFTFSPRHQTVRDCNCSIYPGHVSGHRMAWDMWMWMSPTAALVVSQ
                     LLRIPQAAVDMVAGAHWGVLAGLAYYSMVGNWAKVLVVMLLFAGVDGEPYTTGGTHGR
                     AAHGLTSLFTPGPAQKIQLVVNNGSWHINRTALNCNDSLQTGFLAALFYTHRRNASGC
                     SERMASCRPIDQFDQGWGWPITYNESHGIIDQRPYCWHYAPQPCGIVPALQVCGPVYCFT
                     PSPVVVGTTDRFGAPTYRWGENETDVLLLNNTRPPRGNWFGCTWMNSTGFTKTCGGPF
                     CNIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWL"
ORIGIN
```

Query Match   100.0%; Score 1107; DB 6; Length 1107;
Best Local Similarity   100.0%;
Matches 1107; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 TCCTCCCGCTGCTGGGTAGCGGCCTCACTCCCACGCTCGCGGCCAAGGACGCCAGCATCCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCCTCCCGCTGCTGGGTAGCGGCCTCACTCCCACGCTCGCGGCCAAGGACGCCAGCATCCC 60
```

```
RESULT 2
LOCUS        A32200                                        1107 bp      DNA        linear   PAT 10-DEC-1996
DEFINITION   NANBHV PT polyprotein fragment.
ACCESSION    A32200
VERSION      A32200.1  GI:1926558
KEYWORDS     .
SOURCE       Non-A, non-B hepatitis virus
  ORGANISM   Non-A, non-B hepatitis virus
             Viruses; unclassified viruses.
REFERENCE    1  (bases 1 to 1107)
  AUTHORS
  JOURNAL    Patent: FR 2655990-A 19 21-JUN-1991;
FEATURES             Location/Qualifiers
     source          1..1107
                     /organism="Non-A, non-B hepatitis virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:12440"
                     /clone="136/155"
     CDS             <1..>1107
                     /codon_start=1
                     /product="partial NANBH-PT polyprotein"
                     /protein_id="CAA02092.1"
                     /db_xref="GI:1926559"
                     /translation="SSRCWVALFPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMYVG
                     DLCGSVFLVSQLFTFSPRRHQTVQDCNCSIYPGHVSGHRMAWDMMMNWSPTAALLVSQ
                     LLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLVVMLLFAGVDGEPYTGGTHGR
                     AAHGILTSLFTPGPAQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHRNASGC
                     SERMASCRPIDQFPDQGWGPITYNESHGLDQRPYCWHYAPQPCGIVPALQVCGPVYCFT
                     PSPVVVGTTDRFGAPTYRWGENETDVLLLNNTRPPRGNWFGCTWMNSTGFTKTCGGPP
                     CNIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWL"
ORIGIN
Query Match     100.0%;  Score 1107;  DB 6;  Length 1107;
Best Local Similarity   100.0%;
Matches 1107; Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;
```

```
Qy          1 TCCTCCCGCTGCTGGGTAGCGCCTCACTCCCACGCTCGCGCCCAAGGACGCCAGCATCCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCTCCCGCTGCTGGGTAGCGCCTCACTCCCACGCTCGCGCCCAAGGACGCCAGCATCCCC 60

Qy         61 ACTGCGACAATACGACGCCAGTCGAATTTGCTCGTTCGTTGGGGCGGCTGCCTTCTGCGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACTGCGACAATACGACGCCAGTCGAATTTGCTCGTTCGTTGGGGCGGCTGCCTTCTGCGCT 120

Qy        121 ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTTTCTCAGCTGTTCACCTTCTCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTTTCTCAGCTGTTCACCTTCTCG 180

Qy        181 CCTCGCGACATCAGACGGTACAGGACTGCAATTGTTCAATCTATCCCGGCCCACGTATCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCGCGACATCAGACGGTACAGGACTGCAATTGTTCAATCTATCCCGGCCCACGTATCA 240

Qy        241 GGTCACCGACCATGGCTTGGGGATATGATGATGAAACTGGTCACCTACAGCAGCCTAGTGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGTCACCGACCATGGCTTGGGGATATGATGATGAAACTGGTCACCTACAGCAGCCTAGTGGTA 300

Qy        301 TCGCAGCTACTCCCGGATCCCACAAGCTGTCGTGGACATGGTGGCGGGGGCCCACTGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCGCAGCTACTCCCGGATCCCACAAGCTGTCGTGGACATGGTGGCGGGGGCCCACTGGGGA 360

Qy        361 GTCCTGGCGGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGGTTGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCCTGGCGGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGGTTGTG 420

Qy        421 ATGCTACTCTTTGCCGGCGTTGACGGGGAACCTTACACGACAACCACGGGCCGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGCTACTCTTTGCCGGCGTTGACGGGGAACCTTACACGACAACCACGGGCCGC 480

Qy        481 GCCGCCCACGGGCTTACATCCTCTTCACACCTGGGCCGCGCTCAGAAAATCCAGCTTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCGCCCACGGGCTTACATCCTCTTCACACCTGGGCCGCGCTCAGAAAATCCAGCTTGTA 540
```

RESULT 3
LOCUS   AR144048   1107 bp   DNA   linear   PAT 08-AUG-2001
DEFINITION Sequence 19 from patent US 6210675.
ACCESSION   AR144048
VERSION   AR144048.1 GI:15105915
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1107)
  AUTHORS   Highfield,P.Edmund., Rodgers,B.Colin., Tedder,R.Seton. and
            Barbara,J Anthony.James.
  TITLE   PT-NANB hepatitis polypeptides
  JOURNAL   Patent: US 6210675-A 19 03-APR-2001;
FEATURES   Location/Qualifiers
     source   1..1107
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 1107; DB 6; Length 1107;
  Best Local Similarity   100.0%;
  Matches 1107; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 4
A28155

```
LOCUS           A28155              2116 bp    DNA     linear   PAT 07-JUN-1995
DEFINITION      PT-NANBH mRNA fragment from patent GB2239245.
ACCESSION       A28155
VERSION         A28155.1  GI:1248638
KEYWORDS        .
SOURCE          synthetic construct
  ORGANISM      synthetic construct
                other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2116)
  AUTHORS       .
  TITLE         Post-transfusional non-A non-B hepatitis viral polypeptides
  JOURNAL       Patent: GB 2239245-A 21 26-JUN-1991;
FEATURES             Location/Qualifiers
     source          1..2116
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /note="contig formed by cDNA clones"
                     /db_xref="taxon:32630"
     CDS             308..>2116
                     /codon_start=1
                     /transl_table=11
                     /product="PT-NANBH polyprotein N-terminus"
                     /protein_id="CAA01931.1"
                     /db_xref="GI:1248639"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETAD
                     MIMHTPGCVPCVREGNSSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHQTVQDCNCSIYPGHVSGHRMAWDMMMNWSPTAALVV
                     SQLLRIPQAAVDMVAGAHWGVLAGLAYYSMVGNWAKVLVVMLLFAGVDGEPTTGGTH
                     GRAAHGLTSLFTPGPAQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHRFNAS
                     GCSERMASCRPIDQFDQGWGPITYNESHGLDQRPYCWHYAPQPCGIVPALQVCGPVYC
                     FTPSPVVVGTTDRFGAPTYRWGENETDVLLLNNTRPPRGNWFGCTWMNSTGFTKTCGG
                     PCNIGGVGNNTLICPTDCFRRKHPEATYTKCGSGPWL"
ORIGIN
Query Match           100.0%; Score 1107; DB 6;   Length 2116;
Best Local Similarity 100.0%;
Matches 1107;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy         1 TCCTTCCCGCTGCTGGGATTAGGCTCACTCCCACGCTCCCATCCCCACGCATCCC    60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1010 TCCTTCCCGCTGCTGGGATTAGGCTCACTCCCACGCTCCCATCCCCACGCATCCC  1069

Qy        61 ACTCGGACAAATACGACGCACGTCGATTTGTCGTTCGGTTGGGGCGCGGCTGCCTTCTCGTTCCGCT   120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1070 ACTCGGACAAATACGACGCCACGTCGATTTGTCGTTCGGTTGGGGCGCGGCTGCCTTCTCGTTCCGCT  1129

Qy       121 ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTTTCTCAGCTGTTCACCTTCTCG   180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1130 ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTTTCTCAGCTGTTCACCTTCTCG  1189

Qy       181 CCTCGGCGCCACATCAGACGGTACAGGACTGCAATTGTTCAATCTATCCCGGCCACGTCATCA   240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1190 CCTCGGCGCCACATCAGACGGTACAGGACTGCAATTGTTCAATCTATCCCGGCCACGTCATCA  1249

Qy       241 GGTCACCGCATGGCTTGGGATAATGATGAATGGTCACCTACAACACAGCCCTAGTGGTA   300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1250 GGTCACCGCATGGCTTGGGATAATGATGAATGGTCACCTACAACACAGCCCTAGTGGTA  1309

Qy       301 TCGCAGCTACTCCCGGAATCCCACAAGCTGTGTCGTGTGACATGGTGACGTGGGGA   360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1310 TCGCAGCTACTCCGGAATCCCACAAGCTGTGTCGTGTGACATGGTGACGTGGGGA  1369

Qy       361 GTCCGGCGGCGGCCTTGGCCTACATTGGGGGAACTGGGCTAAGGTCTTGGTTGTGTG   420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1370 GTCCGGCGGCGGCCTTGGCCTACATTGGGGGAACTGGGCTAAGGTCTTGGTTGTGTG  1429

Qy       421 ATGCTACTCTTTGCCCGGCGTTGACGGGGGAACCTTACACGACGAGGGGGACACACCGC   480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1430 ATGCTACTCTTTGCCCGGCGTTGACGGGGGAACCTTACACGACGAGGGGGACACACCGC  1489

Qy       481 GCCGGCCACGGCGTTACATCCCTTCTTCACACCTGGCCTGGGCGCGGCTCAGAAAATCCAGCTTGTA   540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1490 GCCGGCCACGGCGTTACATCCCTTCTTCACACCTGGCCTGGGCGCGGCTCAGAAAATCCAGCTTGTA  1549

Qy       541 AACACCAACGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAATGAC   600
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1550 AACACCAACGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAATGAC  1609

Qy       601 ACTCGGGTTCCTTGCCCGCGTCGTTCTACAGCCACAGGGTTCAGAAATGGGTTCCCGGATTGCTCAGAG   660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1610 ACTCGGGTTCCTTGCCCGCGTCGTTCTACACGCCACAGGGTTCAAATGGGTTCCCGGATTGCTCAGAG  1669

Qy       661 CGCATGGCCAGCTGCCGCGCCCCATTGACCAGTTCGATCAGGGGTCTCCATCACTTAT   720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1670 CGCATGGCCAGCTGCCGCGCCCCATTGACCAGTTCGATCAGGGGTCCATCACTTAT  1729

Qy       721 AATGAGTCCCACGGGCTTGGACCAGAGGCCCTATTTGCTGGACACTCAACCGTGT   780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1730 AATGAGTCCCACGGGCTTGGACCAGAGGCCCTATTTGCTGGACACTCAACCGTGT  1789

Qy       781 GGTATCGTGCCGGGTTGCAGGTTGTGTGGCCCAGTTACTGTTTTCACTCCAAGCCCTGTT   840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1790 GGTATCGTGCCGGGTTGCAGGTTGTGTGGCCCAGTTACTGTTTCACTCCAAGCCCTGTT  1849

Qy       841 GTGGTGGGGACGACGACCGATCGTTTCGGCGCGCCTACGTACAGATGAGGGGTGAGAATGAGACG   900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1850 GTGGTGGGGACGACGACCGATCGTTTCGGCGCGCCTACGTACAGATGAGGGGTGAGAATGAGACG  1909

Qy       901 GACGTGCTGCTTCTCAACAACACACGCCGGCCGCCCCGCGGGCAACTGGTTCGGCGTACATGG   960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1910 GACGTGCTGCTTCTCAACAACACACGCCGGCCGCCCCGCGGGCAACTGGTTCGGCGTACATGG  1969

Qy       961 ATGAAATAGCACCCGGGTTCACCAAGACAGTGTGGGGGGCCCCCCCGTGCAACATCGGGGGGGTC  1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1970 ATGAATAGCACCGGGTTCACCAAGACAGTGTGGGGGGCCCCCCCGTGCAACATCGGGGGGGTC  2029

Qy      1021 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGAAGCATCCGGAGGCCACTTAC  1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2030 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGAAGCATCCGGAGGCCACTTAC  2089

Qy      1081 ACCAAATGCGGTTCGGGGCCTTGGTTG  1107
             |||||||||||||||||||||||||||
Db      2090 ACCAAATGCGGTTCGGGGCCTTGGTTG  2116


RESULT 5
A32202
LOCUS           A32202              2116 bp    DNA     linear   PAT 10-DEC-1996
DEFINITION      NANBHV PT polyprotein fragment.
ACCESSION       A32202
VERSION         A32202.1  GI:1926562
KEYWORDS        .
SOURCE          Non-A, non-B hepatitis virus
  ORGANISM      Non-A, non-B hepatitis virus
                Viruses; unclassified viruses.
REFERENCE   1  (bases 1 to 2116)
  AUTHORS       .
  JOURNAL       Patent: FR 2655990-A 21 21-JUN-1991;
FEATURES             Location/Qualifiers
     source          1..2116
                     /organism="Non-A, non-B hepatitis virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:12440"
     CDS             308..>2116
                     /codon_start=1
                     /product="partial NANBH-PT polyprotein"
                     /protein_id="CAA02094.1"
                     /db_xref="GI:1926563"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYETAD
                     MIMHTPGCVPCVREGNSSRCWVALTPTLAAKDASIPTATIRRHVDLLVGAAAFCSAMY
```

```
VGDLCGSVFLVSQLFTPSPRRHQTVQDCNCSIYPGHVSGHRMAWDMMMWSPTAALVV
SQLLRIPQAVVDMAVAGAHWGVLAGLAYYSVMVGNWAKVLVVMLLFAGVDGEPYTTGGTH
GRAAHGLTSLFTPFGPAQKIQLVNTNGSMHINRTALNCNDSLQTGFLAALFYTHRPNAS
GCSERMASCRPIDQFDQGWGPITYNESHGLDQRPYCWHYAPQPCGIVPALQVCGPVYC
FTPSPVVGTTDRFGAPTYRWGENETDVLLLNNTRPPRGNWFGCTWMNSTGFTKTCG
PPCNIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWL"
```

ORIGIN

```
Query Match             100.0%;  Score 1107;  DB 6;  Length 2116;
Best Local Similarity   100.0%;
Matches 1107;  Conservative   0;  Pred. No. 1e-232;  Mismatches   0;  Indels   0;  Gaps   0;
```

RESULT 6

```
LOCUS       AR144050        2116 bp    DNA    linear    PAT 08-AUG-2001
DEFINITION  Sequence 21 from patent US 6210675.
ACCESSION   AR144050
VERSION     AR144050.1  GI:15105917
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2116)
  AUTHORS   Highfield,P.Edmund., Rodgers,B.Colin., Tedder,R.Seton. and
            Barbara,J.Anthony.James.
  TITLE     PT-NANB hepatitis polypeptides
  JOURNAL   Patent: US 6210675-A 21 03-APR-2001;
FEATURES             Location/Qualifiers
     source          1..2116
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

```
Query Match             100.0%;  Score 1107;  DB 6;  Length 2116;
Best Local Similarity   100.0%;
Matches 1107;  Conservative   0;  Pred. No. 1e-232;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy    481 GCCGCCCACGGGCTTACTCCCTTCTTCACACCTGGGCGTCCGGCCTCAGGAAGGTGTA 540
              ||||||||||||||||||||||||||||||||||||||| ||| ||||| ||||||||
Db   1490 GCCGCCCACGGGCTTACTCCCTTCTTCACACCTGGGCGCCCGGCCCTCAGAAAATCCAGCTTGTA 1549

Qy    541 AACACCAACGGACGGCCACTGAACATCCCTTTGAACTGCCTTGAACTGCCTCCCTCCAA 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 AACACCAACGGACGGCCACTGAACATCCCTTTGAACTGCCTTGAACTGCCTCCCTCCAA 1609

Qy    601 ACTGGGTTCCTTGCGCGGCGTTCTTCTACACGCACCACAGGTTCAATGCGTCCGGATGCTCAGAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 ACTGGGTTCCTTGCGCGGCGTTCTTCTACACGCACCACAGGTTCAATGCGTCCGGATGCTCAGAG 1669

Qy    661 CGCATGGCCAGCTGCCGCCCATTGACCAGTTCGATCAGGGGTCCCATCACTTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 CGCATGGCCAGCTGCCGCCCATTGACCAGTTCGATCAGGGGTCCCATCACTTAT 1729

Qy    721 AATGAGTTCCCACGGCCTTGGACCAGAGGCCCTTATTGCTGGCACTACGCCACTCACCGTGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1730 AATGAGTTCCCACGGCCTTGGACCAGAGGCCCTTATTGCTGGCACTACGCCACTCACCGTGT 1789

Qy    781 GGTATCGTTGGGCGTTGCCGGTTGGTTGGGCCCCAGTGTACTGTTTCACTCCCAAGCCCTGTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1790 GGTATCGTTGGGCGTTGCCGGTTGGTTGGGCCCCAGTGTACTGTTTCACTCCCAAGCCCTGTT 1849

Qy    841 GTGGTTGGGGGACGACCGGATCGTTTCGGCGCCCGCCGTTACGTACACAGGGGTGAGAATGAGACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1850 GTGGTTGGGGGACGACCGGATCGTTTCGGCGCCCGCCGTTACGTACACAGGGGTGAGAATGAGACG 1909

Qy    901 GACGTGCTGCTTCTTCAACAACACGCGGGGCGACCTGGTTGTTGTCGGTGTACATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1910 GACGTGCTGCTTCTTCAACAACACGCGGGGCGACCTGGTTGTTGTCGGTGTACATGG 1969

Qy    961 ATGAATAGACACCGGGGTCACCGGGTTCACCAAGACGTGTGGGGGGGCCACAGCCACTTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1970 ATGAATAGACACCGGGGTCACCGGGTTCACCAAGACGTGTGGGGGGGCCACAGCCACTTG 2029

Qy   1021 GGCAACAACACTTTGATCTGCCCCGAGGGCCACTAC 1080
              ||||||||||||||||||||||||||||||||||||
Db   2030 GGCAACAACACTTTGATCTGCCCCGAGGGCCACTAC 2089

Qy   1081 ACCAAAATGCGGTTCGGGGGCCTTGGTTG 1107
              |||||||||||||||||||||||||||||
Db   2090 ACCAAAATGCGGTTCGGGGGCCTTGGTTG 2116
```

RESULT 7

HPCK1R2

LOCUS   HPCK1R2   9410 bp   RNA   linear   VRL 10-FEB-1999
DEFINITION Hepatitis C virus (strain HCV-1b, clone HCV-K1-R2), complete genome
            sequence.
ACCESSION   D50481
VERSION   D50481.1 GI:1030705
KEYWORDS   polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
            NS5B; envelope protein; non-structural protein;
            IFN-resistant; interferon-sensitive; interferon-resistant;
            IFN-resistant; ISDR; interferon sensitivity determining region;
            HVR; hypervariable region.
SOURCE   Hepatitis C virus
  ORGANISM Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1 (sites)
  AUTHORS   Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
            Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
  TITLE   Comparison of full-length sequences of interferon-sensitive and
            resistant hepatitis C virus 1b. Sensitivity to interferon is
            conferred by amino acid substitutions in the NS5A region
  JOURNAL   J. Clin. Invest. 96 (1), 224-230 (1995)
  MEDLINE   95340824
   PUBMED   7542279
REFERENCE   2 (bases 1 to 9410)
  AUTHORS   Enomoto,N.
  JOURNAL   Unpublished

REFERENCE   3 (bases 1 to 9410)
  AUTHORS   Enomoto,N.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-MAY-1995) Nobuyuki Enomoto, Tokyo Medical and Dental
            University, Second Department of Internal Medicine; 1-5-45 Yushima,
            Bunkyo-ku, Tokyo 113, Japan (E-mail:FXN04522@niftyserve.or.jp,
            Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
FEATURES   Location/Qualifiers
     source   1..9410
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="HCV-1b"
                     /db_xref="taxon:11103"
                     /clone="HCV-K1-R2"
                     /note="IFN-resistant HCV from patient 2"
     CDS   330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAA09072.1"
                     /db_xref="GI:1814086"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNASGVYHVTNDCSNASIYEAAD
                     MIMHTPGCVPCVREANSSRCWVALTPTLAARNSSVPTTTIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALV
                     SQLLRIPQAVMDMVAGAHWGVLAGIAYYSMVGNWAKVLIVMLLFAGVDGNTYVSGGAK
                     SHTTQGLVSLFAPGAQQKIQLVNTNGSWHINRTALNCNDSLNTGFLAALFYTHRFNSS
                     GCPERLASCRSIDAFAQGWGPITYAEPGSSDQRPYCWHYAPRPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRSGAPTYSWGENETDVLLLNNTRAPQGNWFGCTWMNGTGFTKTCGG
                     PPCNIGGVGNNTLICPTDCFRKIPEATYTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PVLSTGLIHLHQNIVDVQYLYGVGSAAVSFVIKWEYILLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVILNAASVAGAHGFLSFLVFFCAAWYIKGRLVPGAAYAFYGWPLLL
                     LLLALPPRAYAMDREMAASCGGAVFVGLALLTLSPHYKVVLARLIWWLQYFITRAEAC
                     LQVWIPPLNVRGGRDAIILLTCALHPELIFSITKILVAILGPLMVLQAGITKMPYFVR
                     AQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHTGLRDLAVAVE
                     PVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEIFLGPADSLEGSGWRLLAPIT
                     AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCTNGVCWTVYHGAGSK
                     TLAGPKGPITQMYTNVDQDLVGWQAPSGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
                     DSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDIPVSEMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSKNVPVEYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGVDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
                     AIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPASGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RAGIYRFVTPGERRSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVWQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTLPLLYRLGAVQNEVILTHPITKYIMACMSADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWQALETFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPRAASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGDMPSTEDLVNLLPAILSPGALVVGVVCAA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLSDFKTWLQSRVLLRLPGVPFFSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNLKCPCQVPAPEFFKELDGVRLHR
                     YAPASKPLLRDEVTFQVGLNQYVVGSQLPCEPEPDVAVLTSMLTDPSHITAEATAKRRL
                     ARGSPPSLASSSASQLSAPSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVVILDSFEPLRAEEDEREVSLPAEILRKSRKFPRAMPIWARDYNPPLIESWKDPDY
                     VPPVVHGCPLPPPRRKRTVVLTESTVSSALAELATKTFGSSESSAADSGT
                     ATAPPDQPSSSDGDPSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEASEGEDVVCCSMS
                     YTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKARLLSVEEACKLTPHSARSKFGYGAKDVRNLSSRAINHI
                     RSWKDLQEDTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
                     VVSTLPQAVMGASYGFQYSPGQRVEFLVNAWKSKKCPMGFAYDTRCFDSTVTENDIRV
                     EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
                     TCYLKATAACRAAKLQDCTMLVCGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPG
                     DPPQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
                     GNIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
                     GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGGRAICGKYL
                     FNWAVRTKLKLTPIPAASQLDLSGWFVAGYSGGDIYHSVSRARPRWFMMCLLLSVGV
                     GIYLLPNR"

```
mat_peptide     330..902
                /product="core protein"
mat_peptide     903..1478
                /product="E1"
mat_peptide     1479..2756?
                /product="E2"
mat_peptide     2757..3407
                /product="E2"
mat_peptide     3408..5300
                /product="NS2"
mat_peptide     5301..5462
                /product="NS3"
mat_peptide     5463..6245
                /product="NS4A"
mat_peptide     6246..7586
                /product="NS4B"
mat_peptide     6246..7586
                /product="NS5A"
mat_peptide     7587..9359
                /product="NS5B"
misc_feature    1479..1559
                /note="hypervariable region (HVR)"
misc_feature    6954..7073
                /note="interferon sensitivity determining region (ISDR)"

ORIGIN

Query Match           85.8%;  Score 950.2;  DB 14;  Length 9410;
Best Local Similarity 91.1%;  Pred. No. 2.4e-198;
Matches 1009;  Conservative  0;  Mismatches 98;  Indels  0;  Gaps  0;

Qy       1 TCCTTCCCGCGTGCTTGGGTAGGCGTCACTCCACTCCCAAGCGCCGAAGGACCCCAGCGATCCCC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1031 TCCTTCCCGCGTGCTTGGGTAGGCGTCACTCCACTCCCAAGCGCCCAGCGATCCCA 1091

Qy      61 ACTGGGACAAATACGGACGCCACAGTCGATTTGCTCGTTGTGGGCGGCTGCCTTTCTGCTTGGGCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1092 ACTACGACAAATACGACGCCACAGTCGATTTGCTCCGTTGGGCGGCTGCCTTTCACTTTCGCTCCGGCT 1151

Qy     121 ATGTACGTGGGGGGATCTCTGCGGATCTGTTTTCCTCAGCTGTTCACCTTCTCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1152 ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTCTGTTCACCTTCTCG 1211

Qy     181 CCTCGCGCCGACATCAGACGGTACAGGGACTGCCAAATTGTTCAATCTTATCCCGGCCACGTATCA 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1212 CCTCGCGCCGGCATCGAGACGGTACAGGGACTGCCAAATTGTTCAATCTTATCCCGGCCACGTATCA 1271

Qy     241 GGTCACCCGCGATGGCTTGGGATATGATGAACTGGTCACCTACAGCAGCCCTAGTGGTTA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1272 GGTCACCCGCGATGGCTTGGGAATATGATGAACTGGTCACCTACAACAGCCCTAGTGGTTA 1331

Qy     301 TGGCAGCTACTCCGGATCCCACAAGCTGTCGTCGTGGACATGGTCGGGGGGCGCCCACTCGGGGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1332 TCGCAGCTACTCCCGGATCCCACAAGCTGTCATCGTGGACATGGTCGGGGGCCCCACTGGGGA 1391

Qy     361 GTCCTGGGCGGGGCCTTGCCTTACTAATTCCATGGTTGGGGGAACTGGCTTAAGGTTCTTGGTTGTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1392 GTCCTGGGCGGGGCCTTGCCTTACTAATTCCATGGTGGGGGGAACTGGCTTAAGGTTCTTGGTTGTG 1451

Qy     421 ATGCTACTCTTTGCCGGCGTTGAGGGGGAACCTTACCACACAGGGGGGGGGGACACACGGGCGGCCGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1452 ATCCTACTCTTTGCCGGTGACGGGGAACACCTACGTGTCACAGGGGGGCGAAGAGCCAT 1511

Qy     481 GCCGCCACAGGGGCGGCTTACATCCCCTCTTCACACCTGGCGGCCTCAGAAAAATCCAGCTTGTA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1512 ACCACCCAAGGGCTTCGGTGTCCCTCTTTGCACCTGGCGGCAGAAAAATCCAGCTTGTG 1571

Qy     541 AACACCAACGGCACGTTGGCACATCAACAGAACTGCCTTGAACTGCAAATGACTCCCTCCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1572 AATACCAACGGCAGTTGGCACATACAAACAGGGACTGCCCGTGAACTGCAAATGACTCCCTCAAC 1631

Qy     601 ACTGGGTTCCTTGCCGCCGCTGTTCCCGCCATGCTTCAGGAG 660
           |||||||||||||||||||||||||||||||||||||||||
Db    1632 ACTGGGTTCCTCGCCGCCGTTGTTCTACACAACAGGTTCAAACTCTTCTGGATGCCCAGAG 1691

Qy     661 CCGCATGGCCAGCTGCCGCGCCCCCATTGACCCAGTTCGATCAGGGGGTGGGGGGTCCCATCACTTAT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1692 CGCCATAGCCAGCTCGCGCACCGCCTCGCCATCGACCGCCATGCGCCCTTCGCCTCAGGGGT 1751

Qy     721 AATGAGTCCCACGGCGCTTGGACCGAGAGGCCCCTTATTGGTTGGGCACTGTGT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1752 GCTGAGCCGGGCAGCTCGGCGGACCAAGAGGCCCCTTATTGGTTGGGCACTGTGT 1811

Qy     781 GGTATCGTGTCGCCGCCGGTTGCAGGTGTGTGGCCCAGTTGTCGCCAGTTCACTCCAAGCCCTGTT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1812 GGTATCGGTACCCCGCCGGAGGTGTGTGGCCAGTGTATTGGTCCACCCCAAGCGCCCGTT 1871

Qy     841 GTGGTGGGGACGGACCGATCGTTTCGGCGCCCCTACGTACAGATGGGGGGTGAGAATGAGACG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1872 GTGGTGGGGACGGACCGATCGTTTCGGCGCCCCTAGGGGGTATAGCTGGGGGAGAATGAGACG 1931

Qy     901 GACGTGCGTGCTTCTCCAACAACACACGCGGCCCCACCACACGGGGCAACTGGTTCGCGCTGTACATGG 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1932 GACGTGCGTGCTTCTCCCTCAACAACACACGCGGGGGGCGCCTCAAGGCAACTGGTTCGCGCTGTACATGG 1991

Qy     961 ATGGAATAAGCACCGGCGGGGGTTCACCAAGACGTGTGGGGGGGGCCCCCCGTGGTCGCAACCATCGGGGGGTC 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1992 ATGGAATGGGCACTGGGTTCACCAAGACGTGCGGGGGGGCCCCCCCCGTGTAACACTCGGGGGGGTC 2051

Qy    1021 GGCAACACACTTTGATCTCGCCCACGGACTGCTTCCGGGGAGGCCACTTAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2052 GGCAACAACACTTTGACCTGCCCACGGACTGCTTCCGGGGAAGCCACACTAC 2111

Qy    1081 ACCAAATGCGGTTCGGGGCCTTGGTTG 1107
           |||||||||||||||||||||||||||
Db    2112 ACCAAATGCGGTTCGGGGCCCTTGGTTG 2138
```

RESULT 8

```
LOCUS       AF313916                9359 bp    mRNA    linear   VRL 01-JAN-2002
DEFINITION  Hepatitis C virus polyprotein mRNA, partial cds.
ACCESSION   AF313916
VERSION     AF313916.1  GI:18027684
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9359)
  AUTHORS   Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
  TITLE     Characteristics of Hepatitis C viral genome associated with disease
            progression in a homogeneous patient population
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9359)
  AUTHORS   Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2000) Medicine, National University of Ireland,
            Cork, Hepatitis C Unit, Clinical Sciences Building, Cork University
            Hospital, Cork, Ireland
FEATURES             Location/Qualifiers
     source          1..9359
                     /organism="Hepatitis C virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11103"
                     /note="isolated from source plasma identified as the
                     single source of the iatrogenic infection of Rhesus
                     negative women in Ireland from May 1977-November 1978;
                     exposure was through HCV 1b contaminated anti-D
                     immunoglobulin"
     5'UTR           1..329
     CDS             330..9359
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAL55821.1"
                     /db_xref="GI:18027685"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED"
```

```
Qy      1 TCCTTCCGGCTTCGTCGGGTAGGCTAGCCCTCCCACCGCTCCTGCTCCGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1032 TCCTTCCGGCTTCGTCGGGTAGCCGTCACTCCCACGCTCCAGCATCCCC 1091

Qy     61 ACTCGGACAATACGACGCACGTCAGTCGATTTCCTCGTTCGGGGCGGGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 ACTACGACAATACGAGTCAGTCGATTTGCTCGTTGGGGACGGGCCGGGCCTAGCCGCCACTGGGGGA 1151

Qy    121 ATGTACGTGGGGGATCTCTGGGGGATCTGTTTCCTTGTCTCTCAGCGTCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152 ATGTACGTGGGGAGATCTNTGCGGGATCTGTTTTCAATCTATCCCGGCGTTTCACCTTCTCG 1211

Qy    181 CCTTCCGGACATCAGGACGGACTCGCAAATTGTTCAATCTATCCCGGCGT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212 CCTTCGCCGGCATGAGACAGTACAGGAGCTGCAAATTGCTCAATCTATCCCGGCCACGTTATCG 1271

Qy    241 GGTCACCGCATGGCTTGGGATATGATGAATGAACTGGTCACCTACAGCAGCCCTAGTGGGTA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

ORIGIN

Best Local Similarity 85.5%;
Matches 1006; Conservative 90.9%;   Pred. No. 2e-197;

Score 946; DB 14; Length 9359;
0; Mismatches 101; Indels 0; Gaps 0;

```
Qy  1272 GGTCACCGCATGGCTTGGGACATGATGAATGAACTGGTCACCTACAGCAGCCCCACTGGGGTG 1331
Qy   301 TCGCAGCTACTCCGGATCCCACAAGCTGTCGTGGACATGGTGGGGAACTGGTGGCCCCACTGGGGA 360
Db  1332 TCGCAGTTACTCCGGATCCCACAAGCTGTCGTGGACATGGTGGGGAACTGGTGGCCCCACTGGGGA 1391
```

ORGANISM Hepatitis C virus
          viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
          Hepacivirus.
REFERENCE 1 (sites)
  AUTHORS Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
           Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
  TITLE   Comparison of full-length sequences of interferon-sensitive and
           resistant hepatitis C virus 1b. Sensitivity to interferon is
           conferred by amino acid substitutions in the NS5A region
  JOURNAL J. Clin. Invest. 96 (1), 224-230 (1995)
  MEDLINE 95340824
   PUBMED 7542279
REFERENCE 2 (bases 1 to 9410)
  AUTHORS Enomoto,N.
  JOURNAL Unpublished
REFERENCE 3 (bases 1 to 9410)
  AUTHORS Enomoto,N.
  TITLE   Direct Submission
  JOURNAL Submitted (08-MAY-1995) Nobuyuki Enomoto, Tokyo Medical and Dental
           University, Second Department of Internal Medicine; 1-5-45 Yushima,
           Bunkyo-ku, Tokyo 113, Japan (E-mail:PXN04522@niftyserve.or.jp,
           Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
FEATURES   Location/Qualifiers
     source   1. .9410
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="HCV-1b"
                     /db_xref="taxon:11103"
                     /clone="HCV-K1-S2"
                     /note="IFN-sensitive HCV from patient 2"
     CDS   330. .9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAA09076.1"
                     /db_xref="GI:1814089"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNASGVYHVTNDCSNASIYEAAD
                     MIMHTPGCVPCVRENNSSRCWVALTPTLAARNSSVPTTTIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRWAWDMMMNWSPTTALVV
                     SQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGRTTVTGGAQ
                     GHTTQRLASLFTPGAQQRIQLVNTNGSWHINRTALNCNDSLNTGFLAALFYAHRFNSS
                     GCPERLASCRSIDAFAQGWGPITYAEPSSSDQRPYCWHYAPRPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRSGAPTYSWGENETDVLLLNNTRAPQGNWFGCTWMNGTGFTKTCGG
                     PPCNIGVGVNNTLTCPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PVLSTGLIHLHQNIVDVQYLYGVSAVVSFVIKWEYILLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVVLNAASVAGAHGFLSFLVFFCAAWYIKGKLVPGAYAFYGVWPLLL
                     LLLALPPRAYAMDREMAASCGGAVFVGLALLTLSPHYKVLARLIWWLQYFITRAEAC
                     LQVWIPPLNVRGGRDAIILLTCALHPELIFSITKILLAILGPLMVLQAGITKMPYFVR
                     AQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVE
                     PVVFSDMETKIITWGADTAACGDIILGLAVSARRGKEIFLGPADSLEGSGWRLLAPIT
                     AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYHGAGSK
                     TLAGSKPITQMYTNVDQDLVGWQAPGGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
                     DSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPEYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGVDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNIGEIPFYGK
                     AIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPASGDVVVVA
                     TDALMGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RAGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVNQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTPLLYRLGAVQNEVILTHPITKYIMACMSAADLEVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWQALETFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPAARAAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGDMPESTEDLVNLLPAILSPGALVVGVVCA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSRVLPRLPGVPFLSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALMRVAAEEYVETRVGDFHYVTGMTTDNLKCPCQVPAPEFFKELDGVRLHR
                     YAPASKPLLRDEVTFQVGLNQYLVVGSQLPCEPEPDVAVLTSMLTDPSHITAERAKRRL
                     (-)

ORIGIN
          Query Match   85.0%;   Score 940.6; DB 14; Length 9410;
          Best Local Similarity 90.6%; Pred. No. 3.1e-196;
          Matches 1003; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy   1 TCCTCCCGCTTGCTGGGTGGGTAGCGCCTCACTCACCGCTTGCTGGGTGGGTAGCGCCTCC   60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1032 TCCTCCCGCGTTGCTGGGTGGGTAGCGCCTCACTCACCGCTTGCTGGGTGGGTAGCGCCTCCA   1091

Qy   61 ACTGCGACAATACGACGCCCAGTCGATTTGCCTGTTGGGGGCTGGGCTGTCCAGCCGTCCA   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 ACTACGACAATACGACGCCCAGTCGACTTGGTCGTTGGGGGCTGGGCTGTCCAGCCGTCCA   1151

Qy   121 ATGTACGTGGGGGATCTCTCCGGATCTGTTGTTTCCCTGGGATCTGTTTCACCTTCTCG   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152 ATGTAGGTGGGGGATCTTCTGGGATCTGTTTTCCCTGGGATCTGTTTCACCTTCTCG   1211

Qy   181 CCTCGCCGCAATCAGACGGTACAGGAGCTGCAATTGTTCAATCTATCCCGGCCACGTATCA   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212 CCTCGCCGCGGCAATGAGACGGTACAGGAGCTGCAATTGTTCAATCTATCCCGGCCACGTTTCA   1271

Qy   241 GGTCACCGCCATGGCTTGGGATATGATGATGAACTGGTCTCCACCTACAGCAGCCCTAGTCGTA   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1272 GGTCACCGCCATGGCTTGGGATATGATGATGAACTGGTCTCCACCTACAGCAGCCCTAGTGGTTA   1331

Qy   301 TCGCAGCTACTCCCGGATCCCACACAAGCTGTCGTGGACATGTGTGGGGGCTGGGCTGTGGGGGA   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 TCGCAGCTACTCCGGGATCCCACAAGCTGTCATGGACATGGTGGCGGGGGCCCACTGGGGA   1391

Qy   361 GTCCTGGCGGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCCTAAGGTGTCCTAAGGTC   420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 GTCCTGGCGGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCCTAAGGTCTTGGTTGTG   1451

Qy   421 ATGCTACTTCTTTTGCCGCGTTGACGGGGAACCTTACACGACAGGGGGGGGACACACGGCCCGC   480
          | ||| ||||||| |||| |||||||| |||||| | |||||||||| ||||||||
Db   - ---   --

                             mat_peptide 330. .902
                                          /product="core protein"
                             mat_peptide 903. .1478
                                          /product="E1"
                             mat_peptide 1479. .2756
                                          /product="E2"
                             mat_peptide 2757. .3407
                                          /product="NS2"
                             mat_peptide 3408. .5300
                                          /product="NS3"
                             mat_peptide 5301. .5462
                                          /product="NS4A"
                             mat_peptide 5463. .6245
                                          /product="NS4B"
                             mat_peptide 6246. .7586
                                          /product="NS5A"
                             mat_peptide 7587. .9359
                                          /product="NS5B"
                             misc_feature 1479. .1559
                                          /note="hypervariable region (HVR)"
                             misc_feature 6954. .7073
                                          /note="interferon sensitivity determinig region (ISDR)"

```
RESULT 10

LOCUS         AY746693              1280 bp    mRNA     linear   VRL 10-OCT-2004
DEFINITION    Hepatitis C virus clone #09-6 polyprotein mRNA, partial cds.
ACCESSION     AY746693
VERSION       AY746693.1  GI:53801768
KEYWORDS      .
SOURCE        Hepatitis C virus
  ORGANISM    Hepatitis C virus
              Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
              Hepacivirus.
REFERENCE     1  (bases 1 to 1280)
  AUTHORS     Chambers,T.J., Fan,X., Droll,D.A., Hembrador,E., Slater,T.,
              Nickells,M.W., Dustin,L.B. and DiBisceglie,A.M.
  TITLE       E1/E2 quasispecies heterogeneity as a pretreatment variable during
              pegylated interferon therapy of chronic HCV infection
  JOURNAL     Unpublished
REFERENCE     2  (bases 1 to 1280)
  AUTHORS     Chambers,T.J., Fan,X., Droll,D.A., Hembrador,E., Slater,T.,
              Nickells,M.W., Dustin,L.B. and DiBisceglie,A.M.
  TITLE       Direct Submission
  JOURNAL     Submitted (10-SEP-2004) Molecular Microbiology and Immunology,
              Saint Louis University School of Medicine, 1402 South Grand Avenue,
              St. Louis, MO 63104, USA
FEATURES             Location/Qualifiers
     source          1..1280
                     /organism="Hepatitis C virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11103"
                     /clone="#09-6"
                     /country="USA"
                     /note="genotype: 1b"
     CDS             <1..>1280
                     /note="contains envelope proteins E1 and E2"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAU94057.1"
                     /db_xref="GI:53801769"
                     /translation="YEVRNASGAYHVTNDCSNSSIVYETMDMIMHTPGCVPREGNS
                     SRCWVALTPTLAARNASIPTTTIRRHVDLLVGAAAFCSMYVGDLCGSVFLVSQLFTF
                     SPRRHETVQDCNCSIYPGHLSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGA
                     HWGVLAGIALYYSMVGNWAKVLIVMLLFAGVDGHTHTTGGSAARNAYGLTTLPTRGPSQ
                     KIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHSFNASGCPERMASCRPINEFAQ
                     GWGPITYTEPHGSDQRPYCWHYAPQPCGTIPAAQVCGPVYCFTPSPVVVGTTDRFGAP
                     TYNWGANETDVLLLNNTRPPQGWMWFGCTWMNGTGFTKTCGGPPCKIGGVGNNTLICPT
                     DCFRKHPEATYTKCGSGPWLTPRCIVDYPYRLWH"
ORIGIN

Query Match        84.4%;  Score 934.2;  DB 14;  Length 1280;
Best Local Similarity   90.2%;  Pred. No. 1e-194;
Matches  999;  Conservative   0;  Mismatches  108;  Indels   0;  Gaps   0;

Qy      1 TCCTCCCGCTGGGTGTAGGGCTCACTCCCACGCTGCGGCCGGCCAGGACGCCAGCATCCCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 TCCTCCTGGCTGCTGGGTAGGGCTCACTCCCACGCGTCGGCCGGCCAGAAACGCCAGCATCCCC 189

Qy     61 ACTGGGACAATACGACGCCACAGTCGAATTGGTCGTTGGGGGGCGGGGGGCCCTTCGCCGCT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 ACTACGACGATACGACGTTCAGGTCCGATTTGCTCGTTGGGGGGGGGGGGGCGCATTCTGCCGTT 249

Qy    121 ATGTTACGGTGGGGGATCTCTCGGGGATCGTGCCAATTGTTCAATTCTATCCCGGCCCGTATCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 ATGTACGTAGGGGGGATCTCTCGGGGAATCGTGCCAACTGCTCCAATCTACCCGGCCATCTTCG 309

Qy    181 CCTCGCCGACATCAGACGGTACAGAGACTGCAATTGTTCAATTCTATCCCGGCCCACGTATCA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 CCTCGCCGCAATGGCTTGGGATTGAGACAGTACAGGATGACGTCCAATCTACCCGGCCATCTATCA 369

Qy    241 GGTCACCGCATGGCTGGGCTTGGGATTAGATGATGAATCGGTCACCTACAGCACGCCCTAGTGGTA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 GGTCACCGCATGGCTTGGGATTGGGATGATGAATGAACTGGTCACCTACAGCACCCTAGTGGTA 429

Qy    301 TCGCAGCTACTCCCGGATCCCACCAAGCTGTCTGTGGACATGGTGGGCGGGGGGCCCACTGGGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TCGCAGTTACTCCGGATCCCACAAGCTGTCTGTGGACATGGTGGGCGGGGGGCCCACTGGGGA 489

Qy    361 CTCCTGGCGGGGCCTTGCCTCACTAATTCCAATGGTGGGGGGGAACTGGGCGTGGGGGGCTAAGG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GTCCTGGCGGGCCTTGCATTACTAATTCCAATGGTGGGGGAACTGGGCGTGGGGGGCTAAGG 549

Qy    421 ATGCTACTCTTTGCCCGGCGTTGACGGGAACCTTACACGACAGGGGGGGGGGGGGGGGGGC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATGCTACTCTTCGCCCGTTGACGGGACACCACCGACAGGGGGGGGGGATCGGCGGCTGC 609

Qy    481 GCCGCCCACCGGGGCTTACATCCCTCCTTCACCACCTGCCGCGGGCTTCAGAAAATCCAGCTTGTA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 AACGCCTACGGGGCTGACGACCCCTCTTCACTCGTGGGGGGTCTCAGAAAATCCAGCTTGTA 669

Qy    541 AACACCAACGGCGAGCTGGCGCATCAACGAGAACTGCCTTGAACTGCAATGACTCCCTCCAA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670 AACACCAACGGCGAGCTGGCACATCAACAGGAACTGCCTTGAACTGCAATGACTCCCTCCAA 729

Qy    601 ACTGGGGTTCCTTGCCCGGCGTGTTCTACACGACAGAGGTTCAATGGCGTCCGGATGCTCAGAG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    730 ACTGGGTTCCTTGCCCGGCTGTTCTATACACATAGCTTCAACGCGTCCGGATGCTCAGAG 789

Qy    661 CGCATGGCCAGTCGCCGCCCCATCAATGAGTTCGCTCACGGGGTGGGGGTTCCCATCACTTAT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    790 CGCATGGCCAGTCGCCGCCCCATCAATGAGTTCGCTCAGGGGTGGGGGCCCATCACTTAC 849
```

```
RESULT 11
LOCUS       AY746700                1280 bp    mRNA    linear   VRL 10-OCT-2004
DEFINITION  Hepatitis C virus clone #09-40 polyprotein mRNA, partial cds.
ACCESSION   AY746700
VERSION     AY746700.1  GI:5801782
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 1280)
  AUTHORS   Chambers,T.J., Fan,X., Droll,D.A., Hembrador,E., Slater,T.,
            Nickells,M.W., Dustin,L.B. and DiBisceglie,A.M.
  TITLE     E1/E2 quasispecies heterogeneity as a pretreatment variable during
            pegylated interferon therapy of chronic HCV infection
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1280)
  AUTHORS   Chambers,T.J., Fan,X., Droll,D.A., Hembrador,E., Slater,T.,
            Nickells,M.W., Dustin,L.B. and DiBisceglie,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2004) Molecular Microbiology and Immunology,
            Saint Louis University School of Medicine, 1402 South Grand Avenue,
            St. Louis, MO 63104, USA
FEATURES             Location/Qualifiers
     source          1..1280
                     /organism="Hepatitis C virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11103"
                     /clone="#09-40"
                     /country="USA"
                     /note="genotype: 1b"
     CDS             <1..>1280
                     /codon_start=1
                     /note="contains envelope proteins E1 and E2"
                     /product="polyprotein"
                     /protein_id="AAU94064.1"
                     /db_xref="GI:5801783"
                     /translation="YEVRNASGAYAVTNDCSNSSIVYETMDMIMHTPGCVPCVREGNS
                     SRCWVALTPTLAARNASIPTTTIRRHVDLLVGAAAFCSVMYVGDLLCGSVFLVSQLFTF
                     SPRRHETVQDCNCSIYPGHLSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGA
                     HWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGHTHTTGGSAARNAYGLTTLFTPGPSQ
                     KIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHSFNASGCPERMASCRPINEPAQ"
```

ORIGIN

```
                GWGPITYTEPHGSDQRPYCWHYAPQPCGTIPAAQVCGPVYCFTPSPVVGTTDRFGAP
                TYNWGANETDVLILNNTRPPQGNWFGCTWMNGTGFTKTCGGPPCKIIGGVGNNTLICPT
                DCFRKHPEATYTKCGSGPWLTPRCIVDYPYRLWH
```

```
Query Match    84.4%;  Score 934.2;  DB 14;  Length 1280;
Best Local Similarity   90.2%;  Pred. No. 1e-194;
Matches  999;  Conservative    0;  Mismatches 108;  Indels    0;  Gaps    0;
```

```
Qy    1 TCCTCCGCCTCTGGGTTAGCCCTCACTCCACTGCCCACGTGCTCGCGCGCCAAGCAGCATCCC 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130 TCCTCTCGCCTGCTGGGTTAGCGCGTGGGTTAGCGGCGGCTAGCGGCGGCCAGAAACGCCAGCATCCC 189

Qy   61 ACTGCCGACAATACGACGGCCAGCAGGTCGGATTTGCTCGGTTCGTTCGTTCTCTGCTCCGCT 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  190 ACTACGACGATTACGACGTCACGTCGGATTTGCTCGGTTCGTTCGTTCCATTCTGCTTCGGTT 249

Qy  121 ATGTACGTTGGGGGGGATCTTCGCCGGATCTGTTTTTCCTGCGTCTTCAGCTTCACCTTCG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  250 ATGTACGTGGGGGGATCTCTGGGGGATCCCTCACTCGTTCTTCCTTGTTTCCCAGCTGTTCACCTTCTCG 309

Qy  181 CCTCGCGCGACATCAGACGGTACAGGACTGCAAATTGTTGTTCAAATTATCCCGGCGTATCA 240
        |||||||||||||||||||||||||||||||||||||||||||||
Db  310 CCTCGCGCGGCATGAGACAGTACAGGACTGCAACTGCTCCATCTACCCCGGCCATCACA 369

Qy  241 GGTCACCGCCATGGCTTGGGATATGATGATGAACTGGTTCACCTTACCAGCTGTGGTA 300
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  370 GGTCACCGCATGGCTTGGGGATATGATGATGAACTGGTCACCACAACAGCCCTTAGTTGGTA 429

Qy  301 TCGCAGCTTACTCCGGGATCCCACACAGAAGCTGTCGTGGGACATGGTGGGCGGGCCCACTGGGGA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  430 TCGCAGTTACTCCGGGATCCCACACAGAAGCTGTCGTGGGACATGGTGGGCGGGCCCCACTGGGGA 489

Qy  361 GTCCTGGCGGGCGCCTTGCCTACTAATTGCATTGTTCATTGGTGGGGAAACTGGGCTTGGTGTG 420
        |||||||||||||  |||||||||||||||||||||||||||
Db  490 GTCCTGGCGGGCGCCTTCGCATACTATTCCATGGTGGGGAACTGGGCTAAGGTTCTGATTGTG 549

Qy  421 ATGCTACTTTTGCCGGGCGTTGACGGGCGGAACCTTACGACGACGGGGGGGGAGCGGCGC 480
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  550 ATGCTACTCTTCGCCGGCGTTGACGGGCACACCCACACGACGACGGGGGGATCGGCGGCGC 609

Qy  481 GCCGGCCCACGGGCGCTTACATCCCTCTTCACACCTGGGCGCGGCTCAGAAAAAATCCAGCTGTA 540
        ||||||||||||  |||||||||||||||||||||||||||
Db  610 AACGCGTTACGGGCCTGACGACGACCCCTCTTCACTCGTGGGCCGGTCTCAGAAAAATCCAGCTTGTA 669

Qy  541 AACACCAAACGCAGCACGTGGGGCACATCAACAGAACAGACTGCCTTGAACTGCAATGACTCTCCCAA 600
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  670 AACACCAAACGGCGAGCTGGCACATCAACAGGACTGCCTTGAACTGCAATGACTCCCTCCAG 729

Qy  601 ACTGGGTTCCTTGCCCGGCGTGTTCTACACGGACCAGGTTCAATGGTCCGGATGGCTCAGAG 660
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  730 ACTGGGTTCCTTGCCCCGGCGTTCTATACACGATAGCTTCAAACGCGTCCGGATGGCCCAGAA 789

Qy  661 CGCATGGCCCAGCTGCCCCGCCCCCATTGACCAGTTCGATCAGGGGGTGGGGTGGGTTCCCATCACTTAT 720
        |||||||||||||||||||||||||||||||||||||||||
Db  790 CGCATGGCCCAGCTGCCCCGCCCCCATCAATGAGTTCGCTCAGGGGGTGGGGTCCCATCACTTAC 849

Qy  721 GGTATCGTGCCCCGGCGTTGCAGGTGTGTGGGCCCAGTGGTACTGTTTTCACTCCAAGCCCTGTT 780
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  850 ACTGAGCCCACGGTTCGGACCAGAGGGCCTTACTTACTGCTGGCACTAGCGCACCTTCAACCGTGT 909

Qy  781 GGTACCACTACCGGGCGACCAGGTTCGGACCAGAGGCCCTTACTCGCTACTAGGACCTACAACAGTT 840
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  910 GGTACCATACCGGGCGACGACCAGGTTCGGACCAGAGGCCCTTACTCGCTACTAGGACCTACAACAGTT 969

Qy  841 GGTATCGTGCCCCGGCGTTGCAGGTGTGTGGGCCCAGTGGTACTGTTTTCACTCCAAGCCCTGTT 900
        |||||||||||||||||||||||||||||||||||||||||
Db  970 GTGGGTGGGGGACGACCAGGTTCGGACCAGAGGCCCAGGTCCACTCAACCGTGT 1029

Qy  901 GACGTGCTGCTTCTCAACAACACACGCGCGCCCCACGCGGGCCAACTGGTTCGGCCTGTTACATGG 960
        ||||||||||||||||||||||||||||||||||||||||||||
Db 1030 GACGTGCTGATTCTCAACAACACACGCGCGCCCACGCGGGCAACTGGTTCGGCCTGTACATGG 1089
```

```
Qy  721 AATGAGTCCCACGGCCTTGGACCAGAGGCCCTATTGCTGGCACTAGCGCACCTTCAACCGTGT 780
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  850 ACTGAGCCCACGGTTGTTCGGACCAGAGGGCCTTACTTACTGCTGGCACTAGCGCACCTTCAACCGTGT 909

Qy  781 GGTATCGTGCCCCGGCGTTGCAGGTGTGTGGGCCCAGTGGTACTGTTTTCACTCCAAGCCCTGTT 840
        |||||||||||||||||||||||||||||||||||||||||||
Db  910 GGTACCATACCGGGCGACCAGGTTCGGACCAGAGGCCCTTACTCGCTACTAGGACCTACAACAGTT 969

Qy  841 GTGGTGGGGACGACCGGATCGTTTCGGCGGCCCCTACGTACAGATGGGGTGAGAATGACG 900
        ||||||||||||||||||||||||||||||||||||||||||||
Db  970 GTGGGTGGGGACGACCGGATCGTTTCGGCGGCCCCTACGTACAGATGGGGTGAGAATGACG 1029

Qy  901 GACCTGCTGCTTCTCAACAACACACGCGCGCCCCACGCGGGCCAACTGGTTCGGCCTGTTACATGG 960
        ||||||||||||||||||||||||||||||||||||||||||||
Db 1030 GACGTGCTGATTCTCAACAACACACGCGCGCCCACGCGGGCAACTGGTTCGGCCTGTACATGG 1089

Qy  961 ATGAATAGCACCGGGGTTCACCCAAGACGTGTGGGGGGCCCCGCCCGTGCAACATCGGGGGTC 1020
        ||||||||||||||||||||||||||||||||||||||||||
Db 1090 ATGGAATGGGACACCGGGTTCACCCAAGACGTGTGGGGGAGCCCCATGTAAAAATCGGGGGTC 1149

Qy 1021 GGCAACACACTTTGATCTGCCCACCACGGACGCATCCCCAGGCGACCCACTTAC 1080
        ||||||||||||||||||||||||||||||||||||||||||
Db 1150 GGCAATAACACCCTGATCTCGCCCACGGACTGCTTCCGGAAGCACCCCGAGGCCACTTAC 1209

Qy 1081 ACCAAATGCGGGTTCGGGCCTTGGGTTG 1107
        ||||||||||||||||||||||||||||
Db 1210 ACCAAGTGCGGGTTCGGGGCCTTGGGTTG 1236
```

RESULT 12
AY746701

```
LOCUS       AY746701                1280 bp    mRNA    linear   VRL 10-OCT-2004
DEFINITION  Hepatitis C virus clone #09-42 polyprotein mRNA, partial cds.
ACCESSION   AY746701
VERSION     AY746701.1  GI:53801784
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 1280)
  AUTHORS   Chambers,T.J., Fan,X., Droll,D.A., Hembrador,E., Slater,T.,
            Nickells,M.W., Dustin,L.B. and DiBisceglie,A.M.
  TITLE     E1/E2 quasispecies heterogeneity as a pretreatment variable during
            pegylated interferon therapy of chronic HCV infection
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1280)
  AUTHORS   Chambers,T.J., Fan,X., Droll,D.A., Hembrador,E., Slater,T.,
            Nickells,M.W., Dustin,L.B. and DiBisceglie,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2004) Molecular Microbiology and Immunology,
            Saint Louis University School of Medicine, 1402 South Grand Avenue,
            St. Louis, MO 63104, USA
FEATURES             Location/Qualifiers
     source          1..1280
                     /organism="Hepatitis C virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11103"
                     /clone="#09-42"
                     /country="USA"
                     /note="genotype: 1b"
     CDS             <1..>1280
                     /codon_start=1
                     /note="contains envelope proteins E1 and E2"
                     /product="polyprotein"
                     /protein_id="AAU94065.1"
                     /db_xref="GI:53801785"
                     /translation="YEVRNASGAYHVTNDCSNSSIVYETMDMIMHTPGCVPCVREGNS
                     SRCWVALTPTLAARNASIPTTTIRRHVDLLVGAAAFCSVMYVGDLLCGSVFLVSQLFTF
                     SPRRHETVQDDCNCSIYPGHLSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGA
                     HWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGHTHTTGGSAARNAYGLTLLFTPGPSQ
                     KIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYTHSFNASGCPERMASCRPINEFAQ
                     GWGPITYTEPHGSDQRPYCWHYAPQPCGTIPAAQVCGPVYCFTPSPVVVGTTDRFGAP
                     TYNWGANETDVLLLNNTRPPQGNWFGCTWMNGTGFTKTCGGPPCKIGGVGNNTLICPT
                     DCFRKHPEATYTKCGSGPWLTPRCIVDYPYRLWH"
ORIGIN
```

Query Match   84.4%;   Score 934.2;   DB 14;   Length 1280;
Best Local Similarity   90.2%;   Pred. No. 1e-194;
Matches   999;   Conservative   0;   Mismatches 108;   Indels   0;   Gaps   0;

```
Qy      1 TCCCTCCCGCTGCTCGGGGTAGCCTCACTCCACCCACGCGGCCAAGGCCGCCACTTAC      60
          ||||| | |||||||| |||||||||||||||||||||||||||||||||||||||
Db     14 TCCCTCTCGGCTGCTCGTTGGGGTAGCGGCTCAGCTAGCGGCGGAAACGCCAGCATCCCC   189

Qy     61 ACTGGGACAATATACGACGCACGTCACGTCGATTTCTTCGTTGTTGGGGCGGGGGGCGCT    120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 ACTACGACGATACGACGAGTCAGTCGATTGCTTCGTTGCTGGGGGCGGGGGGCGCT      249

Qy    961 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGCGGGGGGGGGTC
          ||||| | |||||||||||||||||||||| ||||
Db   1090 ATGAATGGCACCGGGGTTCACCAAGACGTGCGGGAGGGCCCCCATGTAAAATCGGGGGGGGGTC 1020

Qy   1021 GGCAACACACTTTGATCTGCTCCCCACCAGGACTGCTTCCCGAGACGCCACTTAC       1080
          ||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Db   1150 GGCAATAACACCCTGATCTGCCCCACGGACTGCTTCCCGAAAGCCACCCCGGAGGCCACTTAC 1209

Qy   1081 ACCAAATGCGGTTCGGGGGCCTTGGGTTG                                 1107
          ||||| |||| ||||||||||||||||||
Db   1210 ACCAAGTGCGGGTTCGGGGGCCTTGGGTTG                               1236
```

```
Qy    121 ATGTACGTGGGGGGGGATCTTGCGGATCGTTGTTTCCTGCTCTGCTCAGCTTCAGCTTCTCG   180
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 ATGTACGGTGGGGGGGATCTCTGCGGATCGTTGTTTCCTGCTCTGCTCAGCTTCAGCTTCTCG   309

Qy    181 CCTCGCCGACATCAGACGTTACAGGACTGCAATTCTTCAATTCTATCCCGGCCCACGTATCA   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 CCTCGCCCGGCCATGAGACAGCATACAGGACTGCAATTCTTCAATTCTATCCCGGCCATCTATCA 369

Qy    241 GGTCACCGAATGGCTTGGGGATTATGATGATGGAAACTGGTCAACGCTCACTCCAGCTAGTGGTA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 GGTCACCGAATGGCTTGGGGATTATGATGATGGAAACTGGTCAACGCTCACTCCAGCTAGTGGTA 429

Qy    301 TCGCAGCTACTCCGGATCCCACAACCAAGCTGTGTGTGGTGGGGACAATGGTGGTGGGGGCCCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TCGCAGTTACTCCGGATCCCACAACCAAGCTGTGTGGGGGGACAATGGTAGTGGGGCCCCACTGGGA 489

Qy    361 GTCCTGCTGGGGGCCCTTGCCTTACTATTCCCATGGTGGGGAACCTTAAGGGTCTTGGTGTG    420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GTCCTGCTGGGGGCCCTTGCATTACTATTCCCATGGTGGGGAACCTTAAGGGTTCTTGATGTG 549

Qy    421 ATGCTACTCTTTGCCGGCGTCGTTGACGGGCACGGGGGAACCTTACACGCACGAGGGGGGACA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATGCTACTCTTTGCCGGCGTTGACGGACACACCCACACACACGACGACTGGGGACAATGGT 609

Qy    481 GCCGGCCACGGGGCGTTACATCCCTCTTCACACCGTGGGCGCCTCAGAAAATCAGGTTGTA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 AACGCCTACGGGCTGACGACCCTCTTCACTCCGTGGGCGCCTCAGAAAATCAGGTTGTA 669

Qy    541 AACACCAACGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAACGACTCCCCCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670 AACACCAACGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAATGACTCCCTGCAG 729

Qy    601 ACTGGGTTCCTTGCGGCGCTGTTCTACACGCACAGCTTCAATGCTTCGGGATGCTCCAGAG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    730 ACTGGGTTCCTTGCGCTGTTCTATACACATAGCTTCAACGCGTCCGGATGCCCAGAA 789

Qy    661 CGCATGGCCAGCTGCCGCCCCCCATTGAACGAGTTCGATCAGGGGTGGGGTCCCATCTAT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    790 CGCATGGCCAGCTGCCGCCCCCATCAATGAGTTCGCTCAGGGGTGGGGTCCCATCACTTAT 849

Qy    721 AATGAGTCCCACGGGCCGTTGGACCAGAGGCCCCTATTGCTGCACTACCCCACCGTGT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    850 ACTGACCCCACGGGTTCGTTACTCGTGGCACGTGCCCTTACTAGGCGCTTCAACCCGTGT 909

Qy    781 GGTATCGTGCCCGCGCCGTTGCAGGTGTCGTGGCGGCGTTACTGTTTCACTCCAAGCCCTGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    910 GGTACCATACCCGCGCCGGCAGGTGTGCGGTCCGGTGTATTGTTTCACCCCAAGCCCTGTT 969

Qy    841 GTGGTGGGGACGACGGACCGATTCGGTTCGGCCCCCTAGGTACTACAGATGAGGACG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    970 GTGGTGGGGACGACGGACCGATTCGGTTCCGCCCCCTAGTTACTACAGATGAGGACG 1029

Qy    901 GACGTGCTGCTTCTCAACACACGCGCGCCACGGGCAACTGGTTGGTACATGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1030 GACGTGCTGATTCTCAACAACACACGGCCCGCAAGGCAACTGGTTCGGGTGTACATGG 1089

Qy    961 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGCGGGGGGGGGTC 1020
          ||||| | |||||||||||||||||||||| ||||
Db   1090 ATGAATGGCACCGGGGTTCACCAAGACGTGCGGGAGGGCCCCCATGTAAAATCGGGGGGGGGTC 1149

Qy   1021 GGCAACACACTTTGATCTGCTCCCCACCAGGACTGCTTCCCGAGACGCCACTTAC 1080
          ||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Db   1150 GGCAATAACACCCTGATCTGCCCCACGGACTGCTTCCCGAAAGCCACCCCGGAGGCCACTTAC 1209

Qy   1081 ACCAAATGCGGTTCGGGGGCCTTGGGTTG 1107
          ||||| |||| ||||||||||||||||||
Db   1210 ACCAAGTGCGGGTTCGGGGGCCTTGGGTTG 1236
```

RESULT 13
AY545952

```
LOCUS       AY545952                1734 bp    DNA     linear   SYN 28-JUL-2004
DEFINITION  Synthetic construct OH8.11 E1E2 region of HCV polyprotein gene,
```

```
                   complete cds.
ACCESSION   AY545952
VERSION     AY545952.1  GI:46403722
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1734)
  AUTHORS   McKeating,J.A., Zhang,L., Logvinoff,C., Flint,M., Zhang,J., Yu,J.,
            Butera,D., Ho,D.D., Dustin,L.B., Rice,C.M. and Balfe,P.
  TITLE     Diverse Hepatitis C Virus Glycoproteins Mediate Viral Infection in
            a CD81-Dependent Manner
  JOURNAL   J. Virol. 78 (16), 8496-8505 (2004)
   PUBMED   15280458
REFERENCE   2  (bases 1 to 1734)
  AUTHORS   McKeating,J.A., Zhang,L., Logvinoff,C., Flint,M., Zhang,J., Yu,J.,
            Butera,D., Ho,D.D., Dustin,L.B., Rice,C.M. and Balfe,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2004) The Rockefeller University, Center for the
            Study of Hepatitis C, 1230 York Ave, New York, NY 10021, USA
COMMENT     The first 23 and last 36 bases of this set of sequences are from
            the synthetic oligonucleotides used to generate these PCR products.
            Neither the start nor stop codons are naturally present in HCV.
FEATURES             Location/Qualifiers
     source          1..1734
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
                     /clone="OH8.11"
                     /focus
     source          24..1697
                     /organism="Hepatitis C virus"
                     /mol_type="other DNA"
                     /db_xref="taxon:11103"
     CDS             1..1734
                     /codon_start=1
                     /transl_table=11
                     /product="E1E2 region of HCV polyprotein"
                     /protein_id="AAS92923.1"
                     /db_xref="GI:46403723"
                     /translation="MGCSFSIFLLALLSCLTIPASAYEVHNVSGVYHVTNDCSNASIV
                     YEAADLIMHTPGCVPCVRENRSSRCWVALTPTLAARNASVPTTIRRHVDLLVGVAAF
                     CSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPT
                     TALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVIVMLLFAGVDGETYV
                     SGGTAGRNAYGLSTFSVGPSQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYAH
                     RFNSSGCPERMASCRPIDKFDQGWGPITYAEPGSSSDQRPYCWHYAPRPCGIVPAAQVC
                     GPVYCFTPSPVVVGTTDRSGVPTYNWGDNETDVLLLNNTRPPRGNWFGCTWMNSTGFT
                     KTCGGPCNIGGVGNNTLTCPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYP
                     CTVNFTFKVRMYVGGVEHRLTAACNWTRGERCNLEDRDRSELSPLLLSTTEWQVLPC
                     SFTTLPALSTGILHLHQNIVDVQYLYGIGSAVVSFAIKWEYVLLLFLLLADARVCACL
                     WMMLLIAQAEA"
ORIGIN
     primer_bind     1..23
     primer_bind     complement(1698..1734)
```

Query Match   84.4%; Score 934.2; DB 12; Length 1734;
Best Local Similarity   90.2%; Conservative 0; Pred. No. 9.7e-195;
Matches   999; Mismatches 108; Indels 0; Gaps 0;

```
Qy      1 TCCTCCCGCTGCTGGGTAGCGCTCACTCACTCCCAGCGCTCGCGCCGCCAGCAGCATCCCC 60
          |||||||||||||||||||||||||||||||||||||||||||  ||||| |||||||||||
Db    196 TCCTCCCGCTGCTGGGTAGCGCTCGCGCCGCCAGCAGCATCCCC 255

Qy     61 ACTGCGACAATACGACGCCACGTCGGATTTGCTCCTTCGTTGCTCGCCTTCTGCTCGCGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACGACAATACGACGTCATTGCTCTTGTTGTTCCTCGGTGGCAGCTTCTGCTCCGGCT 315

Qy    121 ATGTAGCGTGGGGGGATCTCTCGGATCGTGTTTCACCTTCG 180
          |||||||||||||||||||||||||||||||||||||||||
Db    316 ATGTACGTTGGGGAGATCTCTCGGGATCGTGTTTCACCTTCTCG 375

Qy    181 CCTCGCCCGACATCAGACGGTACAGACTGCAATTCTATCCCGGCCACGTATCA 240
          ||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        CCTCGCCCGACATCAGACGGTACAGACTGCAATTCTATCCCGGCCACGTATCA 240
```

RESULT 14
```
LOCUS       AY545953                1734 bp    DNA     linear   SYN 28-JUL-2004
DEFINITION  Synthetic construct OH8.1 E1E2 region of HCV polyprotein gene,
            complete cds.
ACCESSION   AY545953
VERSION     AY545953.1  GI:46403724
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
Db    376 CCTCGCCGCCATGAGACGGTACAGAGGACTGCAATTGCTCCATCTATCCCGGCCACGTATCA 435
Qy    241 GGTCACCGCCATGGCTTGGGATATGATGATGAATGATGAAACTGGTCACCTACGACAACGGC 300
Db    436 GGCCATACGGCCATGGCTTGGGATATGATGATGAACTGGTCACCTACGACAACGGCCTTAGTGGTG 495
Qy    301 TCGCAGCTACTCCGGATCCCACACAGCTGTCTCGTGGACATGGTGGCGGCGGGGA 360
Db    496 TCGCAGTTGCTCCGGATCCCACAAGCCGTCGTGGGATATGGTGGCGGGCGGGGA 555
Qy    361 GTCCTGGCGGCGCCTTGCCTACTTATTCCATTGGTGGGCGGGCGGGCCCACTGGGGA 420
Db    556 GTCCTAGCGGGCCTTGCCCTACTTATTCCATGTGGGGGCGGGCGGGCCCACTGGGGA 615
Qy    421 ATGCTACTTCTTTGCCCGGCGGTTGCACAGCAGGTTCAATGGGTGTCCGGATGCTCAGAG 480
Db    616 ATGCTACTTCTTTGCCCGGCGGTTGCACAGCAGGTTCAATGGGTGTCCGGATGCTCAGAG 675
Qy    481 GCCGCCCACGGGCCTTACATCCCTCTTCACACCTTGGCCGGGCGGCTCAGAAAATCCAGCT 540
Db    676 AACGCCTTATGGACGCTGTCGACCTTCTTTCAGTTGGGACGCGGGGGTCCCATCACTTAC 735
Qy    541 AACACCAACGCGCAGCTGGGCACATCAACAGGAACTGCCTTGAACTGACAATGACTCCCTCCAA 600
Db    736 AATACCAACGCGCAGCTGGGCACATCAACAGGAACTGCCTCGAACTGACAATGACTCCCTCCAA 795
Qy    601 ACTGGGGTTCCTTGCCCGCGGTGTTCTACAGCGACACGACACAGGTTCAATGCGTCCGGATGCTCAGAG 660
Db    796 ACCGGGTTCCTTGCCCGCGGTGTTCTACGCACACACAGGGGGGCGGCAGGCGC 855
Qy    661 CGCATGGGCCACTGCGGGCCCCCATTGACCCAGTTCGATCAGCAGATCACTTAT 720
Db    856 CGCATGGCCAGCTGCCGGCCCTATCGACAAGTTCGATCAGGGGGTCCCATCACTTAC 915
Qy    721 AATGAGTCCCACACGGCTTGGACAGACGGCCCTTATTGCTGGCACTACGGCAGCCCCGTGT 780
Db        GCTGACCGGGGCAGCTCGGGACTACAGAGCGCCCTTACTGTTGGGCACTACCCGACCGTGC 780
Qy    781 GGTATCGTGCCCGCGGTTGCAGGTGTGGCCAGTTGTACTGTTTCACTCCAGCCCTGTT 840
Db    916 GGTATTGTACCCGCGGGCGGGCCAGGTGTTGGGCCAGTTGTACTGTTTCACTCCAGCCCTGTT 975
Qy    841 GTGGGTGGGGACGACGCGATCGTTTCGCGCGGCCCCTACCTACTACAGCGATGGGAGACG 900
Db    976 GGTTATTGTACCCGGGCCGGGCGCCCCAGGTGTTCCGGGAAAACACCCCGTGGAGAAGAATGAGACG 1035
Qy    901 GACGTGCTGCTTCTCAACAACACGCGGCGCGCCCAACTGGTTCGGGCTGTTCGGGTTACATGG 960
Db   1036 GACGTGCTGCTTCTTCTCAACAACACGCGGCGCCCCAACTGGTTCGGGCTGTTCGGGTTACATGG 1095
Qy    961 ATGAATAGCACCGGGTTCACCAAGACGTGTGGGGGACCCTGCAATATCGGGGGGGTC 1020
Db   1096 ATGAATAGCACTGGGTTCACCAAGACGTGTGGGGGCCCCTGCAACATCGGGGGGGTC 1155
Qy   1021 GGCACAAACACTTTGATCTGCCCCACGGACTGCTTCCGGGAAGCATCCCGAGGCCCACTTAC 1080
Db   1156 ATGAATAGCACTGGGTTCACCAAGACGTGTGGGGGCCCCTGCAACATCGGGGGGGTC 1215
Qy   1081 ACCAAAATGCGGTTCGGGGGCCCTTGGGTTG 1107
Db   1216 GGTAATAACACCCTGACCTGCCCCACGGACTGCTTCCGGAAAACACCCCGAGGCCCACTTAC 1275
Qy                                                                       1302
Db   1276 ACCAAAATGCGGTTCGGGGGCCCTTGGGTTG 1302
```

FEATURES   Location/Qualifiers
     source   1..1734
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
                     /clone="OH8.1"
                     /focus
     source   24..1697
                     /organism="Hepatitis C virus"
                     /mol_type="other DNA"
                     /db_xref="taxon:11103"
     CDS   1..1734
                     /codon_start=1
                     /transl_table=11
                     /product="E1E2 region of HCV polyprotein"
                     /protein_id="AAS92924.1"
                     /db_xref="GI:46403725"
                     /translation="MGCSFSIFLLALLSCLTIPASAYEVEHVNVSGVYHVTNDCSNASIV
                     YEAADLIMHTPGCVPCVRENRSSRCWVALTPTLAARNASVPTTTIRRHVDLLVGVAAF
                     CSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPT
                     TALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGETYV
                     SGGTAGRNAYGLSTFFSVGPSQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYAH
                     RFNSSGCPERMASCRPIDKFDQGWGPITYAEPGSSDQRPYCWHYAPRPCGIVPAAQVC
                     GPVYCFTPSPVVVGTTDRFGVPTYNWGDNETDVLLLNNTRPPRGNWFGCTWMNSTGFT
                     KTCGGPPCNIGGVGNNTLTCPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYP
                     CTVNFTIFKVRMYVAGGVEHRLTAANCNFTRGERCNLEDRDRSELSPLLLSTTEWQVLPC
                     SFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKWEYVLLLFLLLADARVCACL
                     WMMLLIAQAEA"
     ORIGIN
     primer_bind   1..23
     primer_bind   complement(1698..1734)

Query Match   84.4%; Score 934.2; DB 12; Length 1734;
Best Local Similarity   90.2%; Pred. No. 9.7e-195;
Matches 999; Conservative   0; Mismatches 108; Indels   0; Gaps   0;

Qy   1 TCCCTCCCGCTGCTCGGGGTAGCGCTCACTCCCACGCTCCGCCGGGCCAGCCATCCCC 60
          |||||||||| ||| |||||||||||||| |||||||||||||| |||||||||||||||
Db   496 TCCTCCCCGCTGCTCCGGGTAGCGCTCACTCCCACGCTCCGCCGGGCCAGCCATCCCC 555

Qy   61 ACTCGGACAATACGACGACGCACGTCGATTTGCTCGTTGGGGCTTGCCTTCTGCTGCT 120
          |||| ||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   556 GTCCTTAGGCGGCGGCCTGGCCTACTACTATTCCATGGTGGGGAACTGGGCTAAGGTC 615

Qy   121 ATGTAGCTGGGGGGATCTCTCGGGATCTGTTTTCCCGTCTCTCTCAGCGTGTTCACCTTCTCG 180
          |||||||| ||| |||||||||||| || || ||||| |||||||||| |||||||||||||
Db   616 ATGTCACTCTTTTGCCCGTTGGGGCTTGCCTTCTGCTGCT 675

Qy   181 CCTCGGCCGCACATCAGACGGTACAGGACTGCAAATTGTTCAATCTATCCGGCCACGTATCA 240
          |||||||||||||||||||||||||||||||| || |||||||||||||||||||||||||
Db   676 AACGGCCTAGGGGTCGTGACGATGGTTGCTGGAGCTCACTGGGGGGTCCTAGCAGGGCTC 735

Qy   241 GGTCACCGCGCATCAGGGATATGATGATGATGAAACTGGTTCACCTACACCTACGGGGTA 300
          || || |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   736 AATACCAACGGCAGCTGGCACATCAACAGGACTGCCCTGAACTGCAATGACTCCCTCCAA 795

Qy   376 CCTCCCCCGGACATGGCCATGAGACGGTACAGGACTGCAATTGGTCACCTACACCTACGGGTAA 435
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   796 ACCGGGTTCCTTGCCGCGCTGTTCTACGCACAGCGGTTCAATAGCTCGGGCTGCCCAGAG 855

Qy   436 GGCCACATCGGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAACGGCCTTAGTGGTG 495
          ||||||||| |||||| ||| ||||| ||||| ||||||||||||||||||||||||||||
Db   436 GGCCACATCGGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAACGGCCTTAGTGGTG 436

RESULT 15
AY545951
LOCUS   AY545951   1734 bp   DNA   linear   SYN 28-JUL-2004
DEFINITION Synthetic construct OH8.3 E1E2 region of HCV polyprotein gene,
            complete cds.
ACCESSION   AY545951
VERSION   AY545951.1 GI:46403720
KEYWORDS   .
SOURCE   synthetic construct
ORGANISM   synthetic construct
            other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 1734)
AUTHORS   McKeating,J.A.; Zhang,L.; Logvinoff,C.; Flint,M.; Zhang,J.; Yu,J.;
            Butera,D.; Ho,D.D.; Dustin,L.B.; Rice,C.M. and Balfe,P.
TITLE   Diverse Hepatitis C Virus Glycoproteins Mediate Viral Infection in
            a CD81-Dependent Manner

Qy   301 TCGCAGCTACTCCGGATCCCACAAGCTGTTGCTGGATCAGTGGTGGGGGCGCCCACTGGGGGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 TCGCAGGTTGCTTCCGGATCCCACAAGCCGTTGCTGGATATGGTGGGGGCGCCCACTGGGGGA 555

Qy   361 GTCCTTCGGCCGGGCCCTTGCCTACTACTATTCCCAACGGCTTAAGGTCGGGGAACTGGGC 420
          |||||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   556 GTCCTTAGGCGGCGGCCTGGCCTACTACTATTCCATGGTGGGGAACTGGGCTAAGGTC 615

Qy   421 ATGCTACTCTTTTGCCCGGTGTTGACGGGGAACCTTACACGACACAGCGGCACACAGGGCC 480
          |||| ||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   616 ATGTCACTCTTTTGCCCGTTGGGGCTTGCCTTCTGCTGCT 675

Qy   481 GCCGGCCCACGGGCGTTACAATCCCTTCACACCTGGGGGCCGGCCGCTCAGAAAATCCAGTTGTA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   676 AACGGCCTAGGGGTCGTGACGATGGTTGCTGGAGCTCACTGGGGGGTCCTAGCAGGGCTC 735

Qy   541 AACACCAACGGCAGCTGGCACATCAACAGGAACTGCCCTGAACTGCAATGACTCCCTCCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   736 AATACCAACGGCAGCTGGCACATCAACAGGACTGCCCTGAACTGCAATGACTCCCTCCAA 795

Qy   601 ACTGGGGTTCCTTGCCGCGCTGTTCTACGCACAGCGGTTCAATGCGTCCGGATGCCCAGAG 660
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   796 ACCGGGTTCCTTGCCGCGCTGTTCTACGCACAGCGGTTCAATAGCTCGGGCTGCCCAGAG 855

Qy   661 CGCATGGCCAGCTGCCGCCCCCATTGACCAGTTCGATCAGGGGTGGGGCCCATCACTTAT 720
          ||||||| |||||||||| ||| |||||||||||||||||||| |||||||| ||||||
Db   856 CGCATGGCCAGCTGCCGCCCCTATCGACAAGTTCGATCAGGGGTGGGGTCCCATCACTTAC 915

Qy   721 AATGAGTCCCACGGCGTTGGACCAGAGGGCCCTATTCGTGGCACTACCGACCGGTGT 780
          || |||||||||| |||||||||||||| ||||||||||||||||||||||||||||
Db   916 GCTGAGCCGGGGCAGCTCGGACCAGAGGGCCCTACTGTTGGGCACTACCGACCGGTGC 975

Qy   781 GGTATGCGTTGCCCGGCGGTTGCGAGGGTGTGGTGGGCCCCACGGGCTGTTCACTCACGTGTT 840
          || ||||| ||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   976 GGTATTGTACCCGCGGCCAGGGTGTGGCGGGTCCAGTGGTGTTTCACCCCAAGCCCCTGTT 1035

Qy   841 GTGGTGGGGACGACGGATGCGATCGTTTCGGCGCCCCTACAGTTACAGATAGGGGTGAGAATGAGACG 900
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1036 GTGGTGGGGACGACGGATGCCATGGTTTCGGCGTACAACACTGGGGGGGACAAATGAGAGACG 1095

Qy   901 GACGTGCTGCTTCTCAACAACACGGGGCCGCCACCGGGGCAACTGGTTGTTCGGCTGTACATGG 960
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1096 GACGTGCTGCTTCTTAACAACACGAGGCCGCCGCGGGGCAACTGGTTCGGCTGTACATGG 1155

Qy   961 ATGAATAGCACCGGGTTCACCAAGACGTGTGGGGGGCCGCCCTGCAATATCGGGGGGTC 1020
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1156 ATGAATAGCACTGGGTTCACCAAGACGTGCGGGGGGCCCCCGTGCAACATCGGGGGGTC 1215

Qy   1021 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGAAGCATCCGGAGGCCACTTAC 1080
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1216 GGTAATAACACCCTGACCTGCCCCACGGACTGCTTCCGGAAACACCCCGAGGCCACTTAC 1275

Qy   1081 ACCAAAATGCGTTCGGGGCCTTGGTTG 1107
          |||||||||||||||||||||||||||
Db   1276 ACCAAATGCGGTTCGGGGCCTTGGGTG 1302

```
JOURNAL     J. Virol. 78 (16), 8496-8505 (2004)
  PUBMED    15280458
REFERENCE   2 (bases 1 to 1734)
  AUTHORS   McKeating,J.A., Zhang,L., Logvinoff,C., Flint,M., Zhang,J., Yu,J.,
            Butera,D., Ho,D.D., Dustin,L.B., Rice,C.M. and Balfe,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2004) The Rockefeller University, Center for the
            Study of Hepatitis C, 1230 York Ave, New York, NY 10021, USA
COMMENT     The first 23 and last 36 bases of this set of sequences are from
            the synthetic oligonucleotides used to generate these PCR products.
            Neither the start nor stop codons are naturally present in HCV.
FEATURES             Location/Qualifiers
     source          1..1734
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
                     /clone="OH8.3"
                     /focus
     source          24..1697
                     /organism="Hepatitis C virus"
                     /mol_type="other DNA"
                     /db_xref="taxon:11103"
     CDS             1..1734
                     /codon_start=1
                     /transl_table=11
                     /product="E1E2 region of HCV polyprotein"
                     /protein_id="AAS92922.1"
                     /db_xref="GI:46403721"
                     /translation="MGCSFSIFLLALLSCLTIPASAYEVHNNVSGVHVTNDCSNASIV
                     YEAADLIMHTPGCVPCVRENRSSRCWVALTPTLAARNASVPTTTIRRHVDLLVGVAAF
                     CSAMYVGDLCGSVFLVSQLFTFSPRRHETVDQDCNCSIYPGHVSGHRMAWDMMMNWSPT
                     TALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVIVMLLFAGVDGETYV
                     SGGTAGRNAYGLSTFFSVGPSQKIQLVNTNGSWHINRTALNCNDSLQTGFLAALFYAH
                     RFNSSGCPERMASCRPIDKFDQGWGPITYAEPGSSSDQRPYCWHYAPRPCGIVPAAQVC
                     GPVYCFPSXVVVGTTDRFGVPTYNWGDNETDVLLLNNTRPPQGNWFGCTWMNSTGFT
                     KTCGGPPCNIGGVGNNTLTCPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYP
                     CTVNFTIFKVRMVVGGVEHRLTAACNWTRGERCNLEDRDRSELSPLLLSTTEWQVLPC
                     SFTTLPALSTGLIHLHQNIVDVQYLYGISAVVSFAIKWEYVLLFLILADARVCACL
                     WMMLLIAQAEA"
     primer_bind     1..23
     primer_bind     complement(1698..1734)
ORIGIN

Query Match         84.3%;  Score 933.2;  DB 12;  Length 1734;
Best Local Similarity   90.2%;
Matches  998;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;
                                    Pred. No. 1.6e-194;

Qy      1 TCCTCCCGCCTGCTGGGTAGGGCTCACTCCCACGCGCTCGCGCGGCCAAGGACCGCCAGCATCCCC   60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TCCTTCCCGCCTGCTGGGTAGGGCTCACTCCCACGCGCTCGCGCGGCCAAGGACCGCCAGCATCCCC  255

Qy     61 ACTGCGACAAATACGACACGCACACGTCGCGATTTGCTCGATTGCCTTTCTGCTTCCCGCT  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACGACAAATACGACGTCATGTGCATTGCTCGTTGTTGTGGGGTGGCAGCTTTCTGCTTCCCGCT  315

Qy    121 ATGTACGTGGGGGGATCCTGCGGGGATCCTTGTTTTCCTTCCTGCGTTCAGCTGTTCACCTTCTCG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 ATGTACGTGGGGAGGATCTCTGCGGGGATCGTTTTTCCTTGTCTCTCGTCTCTCAGCTGTTCACCTTCTCG  375

Qy    181 CCTCGCCGACATCAGACGGTACAGGACACTGCAAATTGTTCAATCTATCCCGCCACCGTATCA  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 CCTCGCCCGGACCATGAGACGGTACAGGAACTGCAAATTGCTCCATCTATCCCGCCACGGTATCA  435

Qy    241 GGTCACCCGCCATGGCCTTGGGATAATGATGAACTGGTCACACCTACCTACAGCAGCCCCTGTGGTA  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 GGCCATGGCCTTGGGATAATGATGAACTGGTCACCTACACCGCCCTTAGTGGTG  495

Qy    301 TCGGAGCTACTCCGGATCCCACAAGCTGTCGTGTGGACCATTGGCGGGGGCCCCACTGGGGGA  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 TCGGAGTTGCTCCGGATCCCACAAGCGCGTCGTGATATGGTGGGGGCCCCCACTGGGGGA  555

Qy    361 GTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGGAACTGGGGCTTAAGGTCTTGGTTGTG  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    556 GTCCTAGCGGGGCCTTGCCTACTATTCCATGTGGGGAGGGGGAACCTTACACGGACAGGGGGGGGACACACGGCCGC  615

Qy    421 ATGCTACTCTTTGCCGGCGTTGACGGGGAACCTTACACGACAGGAGGGGGGGGACACACGGCCGC  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    616 ATGCTACTCTTTGCCGGCGTTGACGGGGAGAGCCTACACATCAGGGGGGGACGGCCGGC  675

Qy    481 GCCGGCCCACGGGGCCTTACATCCCTCTTCACACACCTGACCCTTCAGAAAATCCAGCTTGTA  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    676 AACGGCCTTATGGGCTGTCGACCTTCTTTTTCAGTTGGGGCGGGTCCCAGAGAAATCCAGCTCGTA  735

Qy    541 AACACCAAGCGGCTGGGCCACATCAACAGAACTGCCTTGAACTGCAATGACTCCCTCCAA  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    736 AATACCAAACGGCCGAGCTGGCCACATCAACAGGAACTGCCCTGAACTGCAATGACTCCCTCCAA  795

Qy    601 ACTGGGGTTCCTTTGCCCGGCGTTGCTCACGCACACAGGTTCAATGCGTCCGGATGCTCAGAG  660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    796 ACCGGGGTTCCTTGCCCGGCGTTTCTACGGCACACAGGTTCAACTCCGTTCCGGATGCCCAGAG  855

Qy    661 CGCATGGCCAGCTGCCGCCCCATTGACCAGTTCGATCAGGGGTGGGGTCCCATCACTTAT  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    856 CGCATGGCCAGCTGCCGCCCCTATCGACAAGTTCGATCAGGGGTGGGGGTCCCATCACTTAC  915

Qy    721 AATGAGTCCCACGGCTTGGACCAGAGACCCTTATTGCTGGCACTAGGGGACTACCACCTCAACCGTGT  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    916 GCTGAGCCCGGGCAGCTCGGACCAGAGACCCTTACTGTTGGCACTACCGACCGTGC  975

Qy    781 GGTATCGTGCCCGGCGTTGCAGGTGTGTGGCCCAGTGGTGACTGTTCACTCCAAGCCCTGTT  840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    976 GGTATTGTACCCGGCGAGGTGCGGTCCAGGTCGTGATTGCTTCACCCCAAGCCNTGTT  1035

Qy    841 GTGGTGGGGACGACGACCGATCGTTTCGGCGCGCCCCTACGTACAGATGGGGTGAGAATGAGACG  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1036 GTGGTGGGGACGACCGATCGTTTCGGCGTCCCCTACGTACAGCTGGGGGGGGACAAATGAGACG  1095

Qy    901 GACGTGCTGCTTCTCAACACACGCGGGGCGCCCAACGGGCAACTGGTTCGGCGTTCTTACATGG  960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1096 GACGTGCTGCTTCTTAACACACGCGGGGGCGCCCAACCTGGTTCGGCGTTACATGG  1155

Qy    961 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGGGCGCCCCCCGTTGCAACATCGGGGGGGGTC  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1156 ATGAATAGCACTGGGGGTTCACCAAGAGACGTGGGGGGGGCGCCCCCGTCAAACATCCCGAGGGTTAC  1215

Qy   1021 GGCACACAAACACTTTGATCTGCCCCACGGGACTGCTTCCGGAAGCAATCCCGAGGGCCACTTAC  1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1216 GGTAATAACACCCTGACCTGCCCCACGGACTGCTTCCGGAAACACCCGGAGGCCACTTAC  1275

Qy   1081 ACCAAAATGCGGTTCGGGGGGCCTTGGGTTG  1107
          |||||||||||||||||||||||||||| ||
Db   1276 ACCAAAATGCGGTTCGGGGGGCCTTGGGCTG  1302
```

Search completed: February 21, 2005, 07:50:41
Job time : 5150 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: February 21, 2005, 04:05:51 ; Search time 702 Seconds
(without alignments)
9334.977 Million cell updates/sec

Title: US-09-664-363-19
Perfect score: 1107
Sequence: 1 TCCTCCCGCTGCTGGGTAGC..........GCGGTTCGGGGGGCCTTGGTTG 1107
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
N_Geneseq_16Dec04:*
1: geneseqnl980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1107	100.0	1107	2	AAQ12240	Aaq12240 Clone 136
2	1107	100.0	2116	2	AAQ12242	Aaq12242 Encodes P
3	927.8	83.8	9605	6	ABK91431	Abk91431 Hepatitis
4	927.8	83.8	9605	6	ABK91424	Abk91424 Hepatitis
5	927.8	83.8	9605	6	ABK91429	Abk91429 Hepatitis
6	927.8	83.8	9605	6	ABK91432	Abk91432 Hepatitis
7	927.8	83.8	9605	6	ABK91411	Abk91411 Hepatitis
8	927.8	83.8	9605	6	ABK91430	Abk91430 Hepatitis
9	927.8	83.8	9605	6	ABK91428	Abk91428 Hepatitis
10	927.8	83.8	9605	6	ABK91425	Abk91425 Hepatitis
11	927.8	83.8	9605	6	ABK91426	Abk91426 Hepatitis
12	927.8	83.8	9605	6	ABK91433	Abk91433 Hepatitis
13	927.8	83.8	9605	6	AAD25332	Aad25332 Hepatitis
14	927.8	83.8	9608	6	ABK91427	Abk91427 Hepatitis
15	927.8	83.8	11062	6	AAD25331	Aad25331 Hepatitis
16	927.8	83.8	11076	3	AAA98965	Aaa98965 Hepatitis
17	919.4	83.1	2187	2	ABA03491	Aba03491 Cuticle p
18	919.4	83.1	2540	2	AAQ43889	Aaq43889 NANB hepa
19	919.4	83.1	2540	2	AAQ63753	Aaq63753 NANBHV ge
20	916.2	82.8	2540	2	AAQ29628	Aaq29628 Hepatitis
21	912.6	82.4	9436	2	AAQ63499	Aaq63499 Blood tra
22	909.8	82.2	9595	2	AAX24843	Aax24843 Infectiou
23	909.8	82.2	9595	4	AAF23492	Aaf23492 Infectiou
24	909.8	82.2	9595	4	AAC86939	Aac86939 Nucleotid
25	909.8	82.2	9595	12	ADO36222	Ado36222 Hepatitis
26	909.8	82.2	9595	12	ADO79396	Ado79396 Hepatitis
27	909.8	82.2	9595	2	AAX24833	Aax24833 Infectiou
28	908.2	82.0	2187	8	ABA03492	Aba03492 Cuticle p
29	904.2	81.7	1953	2	AAL55222	Aal55222 Hepatitis
30	904.2	81.7	2829	2	AAV60673	Aav60673 Plasmid p
31	903.8	81.6	3461	2	AAC86939	Aac86939 Fragment
32	903.2	81.6	3461	2	AAQ64068	Aaq64068 Non-A, no
33	902.2	81.5	2829	2	AAT30386	Aat30386 5'UTR/COR
34	899	81.2	9402	2	ADR82189	Adr82189 Hepatitis
35	897.4	81.1	9405	2	AAQ41345	Aaq41345 Human hep
36	895.8	80.9	9400	13	AAQ40426	Aaq40426 Full-leng
37	893.2	80.7	9414	13	ADR82191	Adr82191 Hepatitis
38	893.2	80.7	9472	2	ADR82190	Adr82190 Korean he
39	892.6	80.6	2238	12	AAQ33282	Aaq33282 Hepatitis
40	892.6	80.6	3081	12	ADJ82982	Adj82982 Adenovect
41	892.6	80.6	37035	12	ADJ82983	Adj82983 Adenovect
42	891	80.5	3360	2	ADJ82985	Adj82985 Adenovect
43	891	80.5	7911	2	AAT03677	Aat03677 Hepatitis
44	891	80.5	8460	12	AAQ32436	Aaq32436 HCV antig
45	891	80.5	9413	2	ADN11771	Adn11771 Hepatitis
					AAQ80498	Aaq80498 DNA encod

ALIGNMENTS

RESULT 1
AAQ12240
ID   AAQ12240 standard; DNA; 1107 BP.
XX
AC   AAQ12240;
XX
DT   25-MAR-2003 (revised)
DT   06-SEP-1991 (first entry)
XX
DE   Clone 136/155 encoding PT-NANBH viral structural protein.
XX
KW   post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
PN   GB2239245-A.
XX
PD   26-JUN-1991.
XX
PF   17-DEC-1990; 90GB-00027250.
XX
PR   17-DEC-1990;
PR   18-DEC-1989; 89GB-00028562.
PR   27-FEB-1990; 90GB-00004414.
PR   03-MAR-1990; 90GB-00004814.
XX
PA   (WELL) WELLCOME FOUND LTD.
PA   (HIGH/) HIGHFIELD P E.
XX
PI   Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ;
XX
DR   WPI; 1991-187584/26.
DR   P-PSDB; AAR12598.
XX
PT   Post-transfusional non-A, non-B hepatitis poly:peptide(e) - and also DNA
PT   and antibodies used in diagnostic assays and in vaccines.
XX
PS   Claim 10; Page 73-76; 108pp; English.
XX
CC   This sequence probably encodes viral structural proteins of the PT-NANBH
CC   viral genome which are antigenic. It was isolated from serum of
CC   infected by the virus. See also Aaq12236-9 and AAq12241-Q12242. (Updated
CC   on 25-MAR-2003 to correct PA field.)

```
XX
SQ  Sequence 1107 BP; 205 A; 344 C; 319 G; 239 T; 0 U; 0 Other;

Query Match      100.0%;  Score 1107;  DB 2;  Length 1107;
Best Local Similarity   100.0%;  Pred. No. 5.6e-308;
Matches 1107;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TCCTCCCGCTGCTGGGTAGCGGCTCACTCCACGCTCGCCAAGGACGGGCATCCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCTCCCGCTGCTGGGTAGCGGCTCACTCCACGCTCGCCAAGGACGGGCATCCCC 60

Qy         61 ACTCGGACAATACAGACGCCACGTCGGATTTGGCTCGTTGGGCGGCGCCTTCTGCTCCGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACTCGGACAATACAGACGCCACGTCGGATTTGGCTCGTTGGGCGGCGCCTTCTGCTCCGCT 120

Qy        121 ATGTACGTGGGGGATCTCTGGGGATCTGTTTCCTCGTTCGTTCACCTTCTCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGTACGTGGGGGATCTCTGCGGGATCTGTTTCCTCGTTCGTTCACCTTCTCG 180

Qy        181 CCTCGCCGACATCAGACGGTACACGACTCGCCAATTGTTCAATATCCGGCCACGTATCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCGCCGACATCAGACGGTACACGACTCGCCAATTGTTCAATATCCCGGCCACGTATCA 240

Qy        241 GGTCACCGCCATGGCTTGGGAATAGTGACTGGTCACCTACAGCACCAGCCTAGTGGTA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGTCACCGCCATGGCTTGGGAATAGTGACTGGTCACCTACAGCACCAGCCCTAGTGGTA 300

Qy        301 TCGCAGCTACTCCGGATCCCACAAGCTGTGTGGACAACATGGTGGGGGGCCCACTGGGGA 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCGCAGCTACTCCGGATCCCACAAGCTGTGTGGACAACATGGTGGGGGGCCCACTGGGGA 360

Qy        361 GTCCTTGGCGCGGCCTTGCCTACTATTCCATGTGGGGACAACTGGGGCTTAAGGTCTTTGGTTGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCCTTGGCGCGGCCTTGCCTACTATTCCATGTGGGGACAACTGGGGCTTAAGGTCTTTGGTTGTG 420

Qy        421 ATGCTACTCTTTGCCGGGCGTTGACGGGGAACCTTACACGACGAGGGGGGGACACACGGGCCCGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGCTACTCTTTGCCGGGCGTTGACGGGGAACCTTACACGACGAGGGGGGGACACACGGGCCCGC 480

Qy        481 GCCGCCCACGGGGCTTACATCCCTCTTTCACACCTGGGCCGGCTCAGAAAAATCCAGCTTGGTA 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCGCCCACGGGGCTTACATCCCTCTTTCACACCTGGGCCGGCTCAGAAAAATCCAGCTTGGTA 540

Qy        541 AACACCAACGGCAGCTGGCACATCAACAGAACTGCCTTGAAACTGCAAATGACTCCCTCCAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AACACCAACGGCAGCTGGCACATCAACAGAACTGCCTTGAAACTGCAAATGACTCCCTCCAA 600

Qy        601 ACTGGGTTCCTTGCCGGGCGTCTTCTTACACGGCCACAGCGTTCAATGGCGTCCGGATGCTCAGAG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACTGGGTTCCTTGCCGGGCGTCTTCTTACACGGCCACAGCGTTCAATGGCGTCCGGATGCTCAGAG 660

Qy        661 CGCATGGGCAGCTGCGGCGCCCCATTGACCAGTTCGATCAGGGGGTGGGGTCCCATCACTTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCATGGGCAGCTGCGGCGCCCCATTGACCAGTTCGATCAGGGGGTGGGGTCCCATCACTTAT 720

Qy        721 AATGAGTCCCACGGCTTGGACCAGAGGCCCTTATTCGTGGCACTACGGCACTCACCGTGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATGAGTCCCACGGCTTGGACCAGAGGCCCTTATTCGTGGCACTACGGCACTCACTTAT 780

Qy        781 GGTATCGTGCCCGCGGTTGCAGGTGTGTGGGCCCAGTGTACTCGTTTCACTCCAAGCCCTGTT 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTATCGTGCCCGCGGTTGCAGGTGTGTGGGCCCAGTGTACTCGTTTCACTCCAAGCCCTGTT 840

Qy        841 GTGGTGGGGACGACGATCGTTTCCGCGCCCTACGTACAGATGGGGGTGAGAATGAGACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGGTGGGGACGACGATCGTTTCCGCGCCCTACGTACAGATGGGGGTGAGAATGAGACG 900

Qy        901 GACGTGCTGCTTCTCAACACACGGGGCCGCACGGGGCAACTGGTTCGGCGTGTACATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACGTGCTGCTTCTCAACACACGGGGCCGCACGGGGCAACTGGTTCGGCGTGTACATGG 960

Qy        961 ATGAATAGCACCGGGGTTCACCAAGAGCGTGTGGGGGGCGCCCCCGTGCAACATCGGGGGGGGGGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGGGCGCCCCCGTGCAACATCGGGGGGGGGGTC 1020

Qy       1021 GGCAACAACACTTTGAATCTGCCCCACCGGACTGCTTCCCGGAAGCANCATCCCGGAGGCCCACTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGCAACAACACTTTGAATCTGCCCCACCGGACTGCTTCCCGGAAGCANCATCCCGGAGGCCCACTTAC 1080

Qy       1081 ACCAAATGCGGTTCGGGGGCCTTGGTTG 1107
              ||||||||||||||||||||||||||||
Db       1081 ACCAAATGCGGTTCGGGGGCCTTGGTTG 1107
```

```
RESULT 2
AAQ12242
ID   AAQ12242 standard; DNA; 2116 BP.
XX
AC   AAQ12242;
XX
DT   25-MAR-2003  (revised)
DT   17-SEP-1991  (first entry)
XX
DE   Encodes PT-NANBH viral structural and non-structural proteins.
XX
KW   post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
FH   Key          Location/Qualifiers
FT   CDS          308..2116
FT                /tag= a
XX
PN   GB2292445-A.
XX
PD   26-JUN-1991.
XX
PF   17-DEC-1990;  90GB-00027250.
XX
PR   18-DEC-1989;  89GB-00028562.
PR   27-FEB-1990;  90GB-00004414.
PR   03-MAR-1990;  90GB-00004814.
XX
PA   (WELL ) WELLCOME FOUND LTD.
PA   (HIGH/) HIGHFIELD P E.
XX
PI   Highfield PE,  Rodgers BC,  Tedder RS,  Barbara JAJ;
XX
DR   WPI; 1991-187584/26.
DR   P-PSDB; AAR12600.
XX
PT   Post-transfusional non-A non-B hepatitis poly:peptide(e) - and also DNA
PT   and antibodies used in diagnostic assays and in vaccines.
XX
PS   Claim 10; Page 83-87; 108pp; English.
XX
CC   This sequence is thought to encode viral structural and non- structural
CC   proteins of the PT-NANBH viral genome which are antigenic. It was
CC   isolated from human serum infectious for the virus. See also AAQ1236-41.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 2116 BP; 650 A; 624 C; 450 T; 0 U; 0 Other;

Query Match      100.0%;  Score 1107;  DB 2;  Length 2116;
Best Local Similarity   100.0%;  Pred. No. 7.2e-308;
Matches 1107;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TCCTCCCGCTGCTGGGTAGCGGCTCACTCCACGCTCGCCAAGGACGGGCATCCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1010 TCCTCCCGCTGCTGGGTAGCGGCTCACTCCACGCTCGCCAAGGACGGGCATCCCC 1069

Qy         61 ACTGCGACAATACGAGCGCCAGTCGATTTGCTCGTTGGGGGGGGGGGGGGGGGCTGCCTTCTGCTCCGCT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1070 ACTGCGACAATACGAGCGCCAGTCGATTTGCTCGTTGGGGGCTGCCTTCTGCTCCGCT 1129
```

```
Qy   121 ATGTACGTGGGGGGGATCCTCTGCGGATCTGTTTTCGTGTTCCTTGTTCTTCAGCTTCTG  180
             |||||||||||||||||    ||||    ||||| |||||||    |  ||||||  | |
Db  1130 ATGTACGTGGGGGGGATCTGTTTTTCGTGCTCTCTCAGCGTGTTCACCTTCTCG      1189

Qy   181 CCTTCGCCGACATCAGACGGTACACGGACACTGCCAATTGTTCAATCTATTCCCGGCTATCA  240
Db  1190 CCTTCGCCGACATCAGACGGTACACGGACACTGCCAATTGTTCAATCTATTCCCGGCTATCA 1249

Qy   241 GGTCACCCGCCGATGGCTTGGGATATCATGATGATGAACTGGTCACCTACAGAGCACCCTAGTGGTA  300
Db  1250 GGTCACCCGCCATGGCTTGGGATATGATGAACTGGTCACCTACAGCAGCCCCTAGTGGTTA 1309

Qy   301 TCGGCAGCTACTCCGGATCCCACAAGCTGTCGTCGTGACATGGTGGGGGGGCCCCACTGGGGA  360
Db  1310 TCGGCAGCTACTCCGGATCCCACAAGCTGTCGTGGACATGGTGGGGGGCCCCACTGGGGA  1369

Qy   361 GTCCTGGCCGGGGCGGGGCCTTGCTTACTATTCCATGGGTGGGCGAACTGGGCTTAAGGTTCTTGGTTGTG  420
Db  1370 GTCCTGGCGGGGGCCTTGCTCTACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGGTTGTG 1429

Qy   421 ATGCTACTCTTTGCCGCGCGTGCTTGACGGGGGAACCTTACACGACACGGGGGACCACCGGCCGCC  480
Db  1430 ATGGCTACTCTTTGCCGGCGCGTTACACGACACCTTACACGGGGGGGACACACGGGCCGC  1489

Qy   481 GCCCGCCCACGGGCGCTTACATCCCTCTTCACACCTGGGCCGCCGGCTCAGAAAAATCCAGCTTGTA  540
Db  1490 GCCCGCCCAGGGGCTTACATCCCTCTTCACACCTGGGCCGCTCAGAAAAATCCAGCTTGTA 1549

Qy   541 AACACCAACGGCAGCTGGGCACATCAACACTGCCTTGAACTGCAATGACTCCCTCCAA  600
Db  1550 AACACCAACGGCAGCTGGGCACATCAACACTGCCTTGAACTGCAATGACTCCCTCCAA  1609

Qy   601 ACTCGGGTTCCTTGCCGCGCTGTTCTACACGGCCACAGGTTCAAATGGCTCCGCGGATGCTTCAGAG  660
Db  1610 ACTCGGGTTCCTTGCCGGCGTGTTCTACACGGCCACAGGTTCAAATGGCTCCGGATGCTTCAGAG 1669

Qy   661 CCGCCATGGCCCAGCTGTTGCGCGCCCCCATTGACAGTTCGATCAGGGGTCCCATCACACTTTAT  720
Db  1670 CCGCCATGGCCAGCTGCCGCCCCCATTGACAGTTCGATCAGGGGGTGGGGTCCCATCACTTTAT 1729

Qy   721 AATGAGTCCCACGGCCTTGGACCAGAGGCCCTATTGCTGGGCACTAGGCACACTCAACCGTGT  780
Db  1730 AATGAGTCCCACGGCCTTGGACCAGAGGCCCTATTGCTGGGCACTAGGCACACTCAACCGTGT 1789

Qy   781 GGTTATCGTGCCCGGCTTGCAGGTGTTGGTGTTACGTGTTCACTCCAAGCCCCTGTT  840
Db  1790 GGTTATCGTGCCCGGGTTGTGGCCCAGTGTACTGTTTCACTCCAAGCCCCTGTT      1849

Qy   841 GTGGTGGGGGAGCGACCGGATACGTTCGGGGCGCCCCCATCCAGTACGATGGGGTGAGAATGAGACG  900
Db  1850 GTGGTGGGGGACGGACCGGATCGTTTCGGGCGCCCCTAGCGTACGATGGGGTGAGAATGAGACG 1909

Qy   901 GAGGTGCTGCTTCTTCAACAACACGCGGGGGCCGGCCGGCAACTGGTTTCGGCGTTGTACATGG  960
Db  1910 GAGGTGCTGCTTCTTCAACAACACGCCGGGGGCCGGCAACTGGTTTCGGCGTTGTACATGG 1969

Qy   961 ATGAATAGCACCCGGCGTTCACCGGGCTGTGGGGGGGGCCCTGCCCTGTGCAACATCGGGGGGGGGTC  1020
Db  1970 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGGGGCCCCCCGTGCAACATCGGGGGGGGGTC 2029

Qy  1021 GGCAACACACTTTGATCTGCTTCCCGGAAGCATCCCGAGGCCACACTTAC  1080
Db  2030 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCCGGAAGCATCCCGAGGCCACACTTAC 2089

Qy  1081 ACCAAATGCGGGTTCGGGGGCCTTGGTTG  1107
Db  2090 ACCAAATGCGGGTTCGGGGGCCTTGGTTG  2116
```

AC   ABK91431;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 8.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX

FH Key	Location/Qualifiers
FH	

FT   CDS   342..9374
FT   /tag= a
FT   /product= "HCV polyprotein"
FT   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation   replace(6934,T)
FT   /tag= b
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cell produced by the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1910 A; 2884 C; 2733 G; 2078 T; 0 U; 0 Other;

RESULT 3
ID   ABK91431 standard; DNA; 9605 BP.
XX

Query Match   83.8%; Score 927.8; DB 6; Length 9605;
Best Local Similarity   89.9%;
Matches 995; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy   1 TCCTCCCGGCGTGCTGGGTAGCGGCTCACTCCCACGCGTCGGGCCGGAGAGCACGCCAGCATCCC 60

```
Db   2124 ACCAAGTGTGGTTCGGGGCCTTGGTTG 2150
```

RESULT 4
ABK91424
ID   ABK91424 standard; DNA; 9605 BP.
XX
AC   ABK91424;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 1.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   342..9374
FT   /*tag= a
FT   /product= "HCV polyprotein"
FT   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation   replace(3625,G)
FT   /*tag= b
XX
PN   WO20025921-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell, the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9

us-09-664-363-19.rng

```
SQ   Sequence 9605 BP; 1910 A; 2884 C; 2732 G; 2079 T; 0 U; 0 Other;

Query Match      83.8%;  Score 927.8;  DB 6;  Length 9605;
Best Local Similarity   89.9%;  Pred. No. 5.1e-256;
Matches  995;  Conservative  0;  Mismatches 112;  Indels  0;  Gaps  0;

Qy        1 TCCTCCCCGTCGTCTGGGGTACCGCTCCTCCCCACGCTTCACTCCCGGCCCCGTGCTGGGCCACGCCCGGCCCAGCANCCC  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1044 TCCTCCCCGGTCGTCTGCTGGGGTAGGGCTCACTCACGCCCCACGCTTCACTCCCGGCCCCGGGCCCAGGGCCAGCANCCC  1103

Qy       61 ACTGCGGACAATACGACGCCCACGTCGATTCGAATTTTGCTCGTTGCTCGCCTTCTTGCTCCGGCT  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1104 ACTACGACGATACGACGCCATGTCGATTTGCTCGTTGCTCGCCTTCTTCTCTCGCCGCT  1163

Qy      121 ATGTACGTGGGGGGATCTTCTCCGGATCGTGTTTTCCTCGTCTTCTTCAGCTTCACCTTCTCG  180
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db     1164 ATGTACGTGGGGAGGATCTCTCGGATCTGCTCGTCGCCTTCGCCTTCACCTTCTCG  1223

Qy      181 CCTCGGCCCGACATCACACGGTTACAGACTGCAAATGTTCAATCTATCATCCCGGCCCAGGTATCA  240
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     1224 CCCTCGCCCGGCCACGAGAGACAGTACAGGACTGCAATTCGCTCAATATTATCCCGGCCCACGTGACA  1283

Qy      241 GGTCACCGGCCATGGCTTGGGGATATGATGAATGAATCGGTCACCTACAGCCCTAGTGGGTA  300
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db     1284 GGTCACCGGCTATGGCTTGGGGGATATGATGAATGAACTGGTCACCTACAGCCAGCCCTAGTGGGTTA  1343

Qy      301 TCGCAGCTACTCTCCGGATCCCACCAAGCTGTCGTGGACATGTGGCGGGGGGCCCACTGGGGA  360
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db     1344 TCGCAGTTACTCCGGATCCCACCAAGCTGTCGTGGGATATGGTTGGGCGGGGGCCCCATTGGGGA  1403

Qy      361 GTCCTTGGGCGGGGGCCCCTACCTACTATTCCCATGGTGGGGAGGGTCTTGGTTGGTG  420
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     1404 GTCCTTAGCGCGGGGCCCTTGCCCTATCCATTGGTGGGGAACTGGGCTTAAGGTTTCTGATTGTG  1463

Qy      421 ATGCTACTCTTTGCCGGCCGTTGACGGGGGGAACCTTACACGACAGGGGGGACACACGGCCCGC  480
            |||||||||||||||||||||||||||||||||||||||||||
Db     1464 ATGCTACTCTTTGCCGGCCGTTGACGGGGGAACCTATGTGACAGGGGGGGGACGATGGCCCAAA  1523

Qy      481 CGCATGGCCAGCTGCCGCGCCCATTGACCAGTTCGATCAGGGGGTCCCATCACTTAT  540
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     1524 AACACCCTCGGGATTTACGTCCCCTCTTTTCACCCGGGGTCATCCCAGAAAATCCAGCTTGTA  1583

Qy      541 AACACCAACGGGCAGCTGGCACATCAAACAGAACTGCCTTGAAACTGGAATGACTCCCTCCAA  600
            |||||||||||||||||||||||||||||||||||||||||||
Db     1584 AACACCAACGGGCAGCTGGCACATCAAACAGACTGCCCTGAACCTGCCAATGACTCCCTCAAC  1643

Qy      601 ACTGGGGTTCCTTGCCGCCGTGTTCTACACGGCCACAGGTTCAAATGCGGTCCGGATGGTTCAGAG  660
            |||||||||||||||||||||||||||||||||||||||||||||
Db     1644 ACTGGGGTTCCTTGCCGCCGTGTTCTACACGGTGCACACTCATCGGGATGCCCAGAG  1703

Qy      661 CGCATGGCCCAGCTGCCGCGCCCATTGACCAGTTCAGGGGGTCCCATCACTTAT  720
            ||| ||||||||||||||||||||||||||||||||||||||||
Db     1704 CGCATGGCCCAGCTCGCAGCGCCCCATCGAACGGTTCGGCTCAGGGGTGGGGGGCCCCATCACTTAC  1763

Qy      721 AATGAGAGTCCCACGCGCTTGACCAGGACCCAGAGGCCCTTATTGTTGGCACTCAGCGGCACCGGTGT  780
            ||||||||||||||||||||||||||||||||||||||||||||
Db     1764 AATGAGTCACACAGCTCCGACACCAGAGGGGCCCTTATCGTGGCACTACGAGCACCCGGCACCGTGC  1823

Qy      781 GGTATCGTACCCCGCGGCCGCTTGCAGGTGGTGTGGGCCCAGTAGTCATGTTTCACTCCAAGCCCTGTT  840
            |||||||||||||||||||||||||||||||||||||||||
Db     1824 GGTATCGTACCCCGGTGCGTCCAGTGTACTTCACCCCCAAGCCCTGTC  1883

Qy      841 GTGGTGGGACAGACCGATCGTTTCGGCGCCCTACGTACAGATGGGGTGAGAATGAGACG  900
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     1884 GTGGTGGGGGGACGACAGCCGGTTGGGGGTGCCTCAGTTGGGGGGCGAGAAATGAGACG  1943

Qy      901 GACGTGCTGCTTCTCAACAACACACGCGGCCCCACGCGGGGGCCAACTGGTTCGGCGTGTACAATGG  960
            ||||||||||||||||||||||||||||||||||||||||||||
Db     1944 GACGGTGCTGCTTTCCTTAACACACGGCCCGGGCCAACTGGTTCGGCTGTTGTACATGG  2003

Qy      961 ATGAAATAGACACCGGGTTCACCAAGAGACGTGGTGTGGGGGGCCCCCGTGTGCAACATCGCGGGGGGGGTC  1020
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     2004 ATGAAATAGACAACTGGGTTCACCAAGAGACGTCGGGCCCCCCGGTGTAACATCGGGGGGGATC  2063

Qy     1021 GGCACAACAACACTTTGATCTGNCCCGCCCCACGGGACTGCTTCCGGAGGGAGCAATCCGGAGGCCACTTAC  1080
            ||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     2064 GGCAATAAAAACCTTGACCTGCCCCCACGGACTGCTTCCGGGAGGCTTCCGGAGGCCAACACCCCACACTTAC  2123

Qy     1081 ACCAAANTGCGGTTCGGGGGCCCTTGGTTG  1107
            ||||| ||||||||||||||||||||||||
Db     2124 ACCAAGTGTGGTTCGGGGGCCCTTGGTTG  2150
```

RESULT 5
ABK91429

ID   ABK91429 standard; DNA; 9605 BP.
XX
AC   ABK91429;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 6.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   342..9374
FT   /*tag= a
FT   /product= "HCV polyprotein"
FT   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation
FT   replace(6859,C)
FT   /*tag= b
XX
PN   WO20259321-A2.
XX
PD   01-AUG-2002.
XX
DR   WPI; 2002-599793/64.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
PS   Claim 9; Page; 69pp; English.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV
PT   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
PT   internal ribosome entry site (IRES) region coding for one or more NS3,
PT   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
PT   are detailed in the specification. Also included are (1) an expression
PT   expression.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and HCV
CC   expression, and HCV replicons useful in studying HCV replication and
CC   proteins, and providing a system for measuring the ability of a compound

```
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1910 A; 2882 C; 2733 G; 2080 T; 0 U; 0 Other;

Query Match         83.8%;  Score 927.8;  DB 6;  Length 9605;
Best Local Similarity   89.9%;  Pred. No. 5.1e-256;
Matches  995;  Conservative   0;  Mismatches 112;  Indels   0;  Gaps   0;

Qy          1  TCCTCCCGCTGCTGGGTAGCGGTCCTTCTGCTCCCACGCTCCACGCTCCCTTCTGCTCCC  60
               |||||||||||||||||||||||||||||||||||   ||||||||||||| |||||||||
Db       1044  TCCTCCCGCTGCTGGGTAGCGGTCCTTCTGCTCCCACGGACCGCCAGCATCCCC  1103

Qy         61  ACTCGGACAATACGACACGCCACGTCGGATTCGAATTTGCTCACTCCTTCTGCTCCGCT  120
               ||| ||||| || ||||||||| ||  |||||||||||| |||  ||||||| |||||||
Db       1104  ACTACGACGATACGACGCCATGTCGATTTTGCTCGTTGTGGCGTCTCTGCTCCGCT  1163

Qy        121  ATGTTACGTGGGGGATCTTCTGCGGATCTGTTTTTCCTCGTTGTTGTTCTCCTTCTCG  180
               ||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db       1164  ATGTTACGTGGGGGATCTCTGCGGATCTGTTTTTCCTCGTTGTGGCGTCTCTGCTCCGCT  1223

Qy        181  CCTCGCCGACATCAGACGGTACAGACGACTCGCAAATTGTTCAATCTATCCCGCCACTATCA  240
               |||||||||| |||||||||||||||||||||||||||| |||||||  |||||||||||||
Db       1224  CCTCGCCGCGAGACAGTACAGACGACTGCAATTGCTCAATATATCCCGCCACGTGACA  1283

Qy        241  GGTCACGCGCATGGGGTCTTGGGATGATGATGATGAACTGTTCACCTACAGCAGCCTAGTGGTTA  300
               |||||||| |||||| |||||  |||| ||||| ||| ||||| |||| |||||||||||||||
Db       1284  GGTCACGGTATGGGCTTGGGATATGATGATGAACTGGTCACCTACAGCAGCCCTAGTGGTTA  1343

Qy        301  TCGCAGCTACTCCGGATCCCACCAACGCTGTGCGTGGACATGGTGCGGGGCGGGCCCACTGGGGA  360
               ||||||||||||  ||||||||||||||| ||||||||||||||||| |||||||||||||||
Db       1344  TCGGAGTTACTCCGGATCCCACCAAGCTGTCGTCGTGGGTGGATATGGCGGGGGGCCCATTGGGGA  1403

Qy        361  GTCCTGGCGGCGGGCCTTGCCTACTACTATTCCATGGGGGGGAACTGGGCTAAGGTCTTGGTTGTTG  420
               |||||||||||||||| |||||||||||||||||||| |||||||||||||||| |||||| ||||
Db       1404  GTCCTTAGCGGCGGGCCTTGCCTACTATTCCATGGTGGGGGCTAAGGTTCTGGTTCTGGATTGTGTG  1463

Qy        421  ATGCTACTCTTTGCCGCGGGTTGACGGGGAACCTTACACGACGAGGGGGGGGGACACGGCCGCGC  480
               |||||||||||||||| |||  | ||||||||| |||||| |||  |||||||||||||||||||
Db       1464  ATGCTACTCTTTGCCGCGGGTTGACGGGGAACCTATGTGACGAGTTGACGGATGCCAAA  1523

Qy        481  GCCGCCCACCGGGGCCTTACACTCCCCCATTGACCACAGTTCAGATCAGGGGTGGGGTCCCAGGGCCTTGTA  540
               ||||||| |||||| |||||||||||| |||||| ||||||||||||||||||||||| |||||||| ||||
Db       1524  AACACCTCGGGATTACGTCCCCTCTTTTCACCGGGTCATCCCAGAGAAATCCAGCTTGTA  1583

Qy        541  AACACCAACGGCGACGTGGCACATCAACAGACTGCCTTGAACTGAAATGACTCCTCCAA  600
               |||||||| |||||||| |||||||||||||||| ||| ||||||||| ||||||||||
Db       1584  AACACCAACGGCAGCGTGGGACATCACAGACTGCGAAATGACTGCCTCCTCAC  1643

Qy        601  ACTGGGGTTCCTTGCGCGGCTGTTCTACACGGACAAGGTTCCGGAGCTGCTCAGAG  660
               |||||||||||||||| ||||||||| ||| ||||||||||| |||| |||||||
Db       1644  ACTCGGGTTCCTTGCTCGGCTGTTCTACGTGCACAAGCTTCAACTCTGGATGCCCAGAG  1703

Qy        661  CGCATGGCCAGCTGCCGCGCCCATTGACCACAGTTCGAATCAGGGGTGGGGTCCCATCACTTAT  720
               ||||||| |||||||||||||||||||||| |||| |||||||||||||||  ||||||||||||
Db       1704  CGCATGGCCAGCTGCAGCCCCATCGACGCGGTCGCTCAGGGCGGTGGGGTCCCATCACTTAC  1763

Qy        721  AATGAGTCCCACGGCCTTGGCACTACGCCACTCACCGGTGT  780
               |||| |||||||||||| ||||||| ||||||| ||||||
Db       1764  AATAAGTCACCAGCTGCCCTATGTGGCCACTACGCACCCGCCGTGC  1823

Qy        781  GGTATGCGTGCCGCGCGGTTGCAGGTGTGGGCCCCAGTGTATTGTTCACTCACAGCCCTGTT  840
               |||| |||||| |||||||| |||| ||| ||||||||| |||||| ||||||||||||||
Db       1824  GGTATCGTACCGCGGCGGTGGTCTGGGTGGTCACCGTCTTCACCCAGACGCCCTGTC  1883

Qy        841  GTGGTGGGGACGACGACCGGATTCTTTCCGCGGCGCCCTAGGTAACAGATGGGGTGAGAATGAGACG  900
               |||||||||| |||||||||||||||||||| || |||||||| ||||||||||| |||||||||||
Db       1884  GTGGTGGGGACGACGACCGACCGGTTCGGCGTCCTACGTACAGTTGGGGGGAGAATGAGACG  1943

Qy        901  GACGTGCTGCTTCTCAACAACACACGGCGGCCGCCCACCGGGGCAACTGGTTCGGGCTGTACATGG  960
               |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1944  GACGTGCTGCTTCTTAACAACAGACGGCGGCCGCCCACCGGGGCAACTGGTTCGGGCTGTACATGG  2003

Qy        961  ATGAATAGCACCGGGTTCACCAAGACGACGTGGTGGGCGCCCCGCCCCCCGCAAGGCAAACATGGG  1020
               ||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       2004  ATGAATAGCACTGGGTTCACCAAGACGTGGGGGCCCCCCGCCCCCCGCAAGGCAAACATGGG  2063

Qy       1021  GGCAACAACACTTTGATCTCGCCCCACGGACTGCTTCCGGGGAAGCATCCCGGAGCCCACTTAC  1080
               |||||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||||
Db       2064  GGCAATAAAACCTTGACCCTGCCCCACGGACTGCTTCCGGGGAAGCACCCGGAGCCCACTTAC  2123

Qy       1081  ACCAAATGCGGTTCGGGGCCCTTGGTTG  1107
               |||||| ||||||||||||||||||||||
Db       2124  ACCAAGTGTGGTTCGGGGCCCTTGGTTG  2150
```

CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC replicon enhanced cell or which containing a functional HCV replicon; (5)
CC an HCV replicon enhanced cells made in the method; and (6) measuring the
CC ability of a compound to affect HCV activity. The HCV replicons and HCV
CC replicon enhanced cells are useful in studying HCV replication and HCV
CC expression, and host, cell interactions, producing HCV RNA and
CC proteins, and providing a system for measuring the ability of a compound
CC to modulate one or more HCV activities e.g. to discover drugs which may
CC treat HCV mediated diseases such as liver failure, cirrhosis and
CC hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC mutant of the invention. Note: The present sequence is not shown in the
CC specification but was created by the indexer using the HCV sequence
CC appearing as ABK91411 and the information in Claim 9
XX
SQ Sequence 9605 BP; 1911 A; 2883 C; 2732 G; 2079 T; 0 U; 0 Other;

Query Match   83.8%; Score 927.8; DB 6; Length 9605;
Best Local Similarity 89.9%;
Matches 995; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy   1 TCCTCCCGCTGCTGGGGTAGCCCTCACTCCCACGCTCGCCGGCCGCCAAGGACGCCAGCATCCCC   60
           |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   1044 TCCTCCCGCTGCTGGGGTAGGGCTCACTCACTCCCACGCTCGCCGGCCGCCAAGGACGCTAGCGTCCCC 1103

Qy   61 ACTCCGACAAATACGACGCCACGTCGTCCATTTTCCTCGTTCGTTGGTGGCGTGCTTCGTGCTCCGCT   120
           |||| ||| || |||| ||||||||| |||||||||||||||||||||| || |||| |||||||||
Db   1104 ACTACGACGATACGACGCCCATGTCGATTTGCTCGTTTGGGGCGGCGTGCTTCGTGCTCTCTGCTTCCGGCT 1163

Qy   121 ATGTACGTGCGGGGATCTCGCGGATCGTAGTTTTTCCGTCGTTCACCTTCTCG   180
           ||| |||| |||||| |||||| |||||||| || |||||||||| ||||||||
Db   1164 ATGTACGTGGGGAGATCTCTGCGGACGTCGTTTTTCCGTCGTTCACCTTCTCG 1223

Qy   181 CCTCGCCCGACATCAGACGGTACAGGACTGCAAATTGTTCAATCATATCCCGGCCAGCCTATCA   240
           ||| || || ||| |||||| |||| |||| |||||||||||||||||||| |||||| || |||
Db   1224 CCTCGCCCACGACGAGACAGTACAGGACTGCGAATTGCTCAATATATCCCGGCCACGTGACA 1283

Qy   241 GGTCACCGCGATGGCTCTTGGGATATGTAATGAATCGGTCACCTACACGCCCTAGTGGTTA   300
           |||||| |||||| ||||| |||| |||| |||||| |||| |||||| |||||| || |||
Db   1284 GGTCACCGTATGGGCTTGGGATAATGAATCGGTCACCTACAGCAGCCCTAGTGGTA 1343

Qy   301 TCGGCAGCTACTCCGGATCCCAACAAGCTGTCGTGGACATGGTGGCGGGGGGCCACTGGGGA   360
           ||||| ||||||| ||||||||||| |||| |||| ||||| ||||||||||| ||||||||
Db   1344 TCGGCAGTTACTCCGGATCCCACACGGTGTCGTGGAATGGGGCCCCATTGGGGA 1403

Qy   361 GTCCTGGCGCGGGCCTTGCCCTACTATTCCATGGTGGGGAACTGGGCTTAAGGTCTTGGTTGTGTG   420
           ||||||||||||| |||| |||| |||||||||||||| |||||| |||| ||||||| ||||||||
Db   1404 GTCCTAGCGGCGGGCCTTGCCCTACTATTCCATGGTGGGGAACTGGGGTTAAGGTTTCGAATGTGTG 1463

Qy   421 ATGCTACTCTTTGCCGCGGCGTTGACGGGGGAACCTTACACGACACAGGGGACACACGGCCGC   480
           |||| ||||||| |||| ||||| ||||||||||||||| |||||||||||| |||||||||||
Db   1464 ATGTACTCTTTGCCGGGCGTTGACGGGGGAACCTATGTGACAGGGGAACCGCGACGGATGGCCAA 1523

Qy   481 GCCGCCCACGGGGCTTACATCCCTCTTCACACCTGGGCCGCGGCTCAGAGAAATCAGCTTGTA   540
           |||| |||||||| ||||||||||| |||||| ||||||||||| ||||| ||||||||||||
Db   1524 AACACCCTCGGGGATTACGTCCCCTCTTTTCACCCGGGGTCATCCCAGAAAATTCAGCTTGTTA 1583

Qy   541 AACACCCACGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAATGACTCCCTCCAA   600
           |||||| ||||||||||||||| |||||||| |||||| ||||| |||||||| |||||||
Db   1584 AACACCAACGGCAGCTGGCACATCAACAGGACTGCCCTGAACTGCAATGACTCCCTCAAC 1643

Qy   601 ACTGGGGTTCCTTGCCCGCGCGTGTTCTTCACAGCACAGGTTCTCCGGATAGTGCTTCGAGAG   660
           ||||||||||||||||||| ||||| |||||| | |||||||| |||| || ||||||||||
Db   1644 ACTCGGGTTCCTTGCTGCGCGTTCTCACTCTCGACGATGCCCAGAG 1703

Qy   661 CGCATGGCCAGCTGCCGCCCCATTGACCAGTTCAGGGTGGGTCCCCATACTGT   720
           || |||||||||||||| |||| |||||||||| ||||| |||||| |||||||
Db   1704 CGCATAGGCCAGCTCGCAGCCCCATGCGGCGGGTGGGGGGCCCCATCACTCAC 1763

Qy   721 AATGAGTCCCACGGCTTGGACCAGAGGCCCTATTCCTGCACCTACGGACACGGTGT   780
           ||||| |||||| ||||| |||||||| ||||| |||| |||| ||||||||||||
Db   1764 AATGAGTCACAGCTCGGACCAGAGGCCCTTATTGTTGGGCTACAGGCCACCCTGC 1823

Qy   781 GGTATCGTGCCCCGCGGTTGCAGGGTGTGGCCCAGTGTACTGGTTCACTCCAAGCCCTGTTT   840
           |||| |||||||| |||||| |||||| || |||||| |||| |||||| |||||||||||
Db   1824 GGTATCGTACCCGGCGGGCGGCGGCCGCGGGCGCAGGGCGTGGGTCAGGTGGTCTCAGCTGCCGCC 1883

Qy   841 ...   900
Db   1884 GTGGTGGGGGACGAACCGGATCGTGTTCGCGGGCGTGGGGGCGGCGGCGGCGGCGGCGGCCGCCG 1943

Qy   901 ...   960
Db   1944 ... 2003

Qy   961 ... 1020
Db   2004 ... 2063

Qy   1021 ... 1080
Db   2064 ... 2123

Qy   1081 ACCAAATGCGGTTCGGGGGGCCTTGGGTTG 1107
           ||||| ||||||||||||||||||||||||
Db   2124 ACCAAGTGTGGTTCGGGGGGCCTTGGGTTG 2150

RESULT 7
ABK91411
ID   ABK91411 standard; DNA; 9605 BP.
XX
AC   ABK91411;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication.
XX
OS   Hepatitis C virus.
XX
FH   Key   Location/Qualifiers
FT   CDS   342..9374
FT   /*tag= a
FT   /product= "HCV polyprotein"
FT   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
DR   P-PSDB; ABG32451.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page 36-39; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations

```
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and HCV
CC   expression, and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is the HCV replicon Con 1,
CC   used as a basis for the adaptive mutations of the invention
XX
SQ   Sequence 9605 BP; 1910 A; 2883 C; 2733 G; 2079 T; 0 U; 0 Other;

Query Match          83.8%;  Score 927.8;  DB 6;  Length 9605;
Best Local Similarity 89.9%;  Pred. No. 5.1e-256;
Matches 995;  Conservative 0;  Mismatches 112;  Indels 0;  Gaps 0;

Qy          1 TCCTTCCCGCTTGGTCTGGGTAGGCGTCCTCACTCCCACGCTCGCGGCCCAAGCAGGACGCCCAGCATCCCC   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1044 TCCTTCCCGCTTGGTCTGGGTAGGCGTCCTCACTCCCACGCTCGCGGCCCAAGCAGGACGCCCAGCATCCCC 1103

Qy         61 ACTCGGACAAATACGACGCCACGTCGATTTCGCTCGTTCGGGGCGGGTGCCTTCTGGTCCCGCT          120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ACTACGACGAATACGACACGGCCATGTCGAATTTGCTCGTTCGGGGCGGGTGCCTTCTGGTCCCGCT       1163

Qy        121 ATGTACGTGGGGGATCTCTGCGGATCTGTTTTTCCTGCGGGATCTTGTTTTCCTGTGCCCAGCTGTTCACCTTCTCG   180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1164 ATGTACGTGGGGAGATCTCTGCGGGATCTGTTTTTCCTGCGGGATCTTGTTTTCCTGTGCCCAGCTGTTCACCTTCTCG 1223

Qy        181 CCTCGGCCCGACATCAGGACGGTACAGGAGGACCTGCACAATTGTTCAATCTATTCCCGGCCCACGTATCA   240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 CCTCGGCCCGGCACGGACGACAGTACAGGAGGACTGCGAATTGCTCAATATATCCCGGCCCACGTGACA     1283

Qy        241 GGTCACCGCCATGGGTTGGGGATAGTGATGATGACTGGTCACCTACAGCAGCCCCTAGTGGTA         300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 GGTCACCGTTATGGCTTGGGGATAATGATGAACTGGTCACCTACAGCAGCCCTAGTGGTTA       1343

Qy        301 TCGCGCAGCTACTCCGGATCCCACAAGCTGTCGTGGACATGCTGGGGCGGGGCCCCACTGGGGGGA   360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TCGCAGTTACTCCGGATCCCACAAGCTGTCGTGGATATGGTGGGGCGGGGCCCCATTGGGGA        1403

Qy        361 GTCCTGGGCGGGGCCTTGCCTTACTATTCCATGGTGGGGAACTCGGCTAAGGTCTTGGTTGGTGTG   420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 GTCCTAGCGGGGCCTTGACCTTACTATTCCATGGTGGGGAACTGGGCTAAGGTTCTGATTGTG       1463

Qy        421 ATGCTACTCTTTGCCGGCGGTTGACGGGGGACCTTACACGACGAGGGGGGACACCGGCCCGC   480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ATGCTACTCTTTTGCCCGCGGTTGACGGGGGAACCTTACGACGAGGGGGGACGATGGCCAAA 1523

Qy        481 GCCCGCCACGGGCTTACACTCCCTCTTCACCTGGCCGCCGGGCTCAGAAAAATCCAGCTTGTA   540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AACACCAACGGCAGCTGGCACATCAACAGGACTGCCCTCGAAATGACTCCCTCAAC         1583

Qy        541 AACACCAACGGCAGCTGGCACATCAACAGGAACTGCCTTGAACTGCAAATGACTCCCTCCAA   600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 AACACCAACGGCAGCTGGCACATCAACAGGACTGCCCTCGAAATGACTCCCTCAAC         1643

Qy        601 ACTCGGGTTCCTTGCCGCGCCCCATGAACCAGTTCAAATGCGTCCGGATGTCTCCGGATAGCCAGAG   660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 ACTCGGGTTCCTTGCTTGCTCGGTGTCTACACTCATCTGGATGCCAGAG                    1703

Qy        661 CGCAATGGCCCAGCTGCGCGCCCCATTGAACCAGTTCGAATCAGGGGGTGGGGGTCCCATCATCTTAT   720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 CGCATAGGCCAGCTGCGAGCCCCATCGACGCGCCCTATTCGCTCAGGGGGTGGGGGCCCATCACTTAC 1763

Qy        721 AATGAGTCCCACGGGCTTGGACCAGAGAGGCCCTTATTGCTGGCACTACGGCCACTCCAACCGTGT   780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 ...
```

```
Db       1764 AATGAGTCACACAGTCTCGGACCAGAGAGGCCTTATTGTTGGCACTACGGCCACCCCGGGCCGTGC   1823
Qy        781 GGTATCGTGGCCGCCGGTTGCAGGGTGTGGGGTCGGGCCCCAGTGTACGTGTTCACTCCAAGCCCTGTT   840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1824 GGTATCGTGTACCCGGGCGGCGCAGGTGTGTGTGGTTCCAGTGTGTTCATTCCAAGCCCTGTC        1883

Qy        841 GTGGTGGGGACGGACCGATCGTTTCGGCGCCCCTACGTACAGAGATGGGGGTGAGAATGAGACG   900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1884 GTGGTGGGGACGGACCGGTTCGGCGTCCCTACGGTACAGTTGGGGGTGAGAATGAGACG        1943

Qy        901 GACGGTGCTGCTTCTCAACACACCGGGCGCGCCCGCCACCGGGCAACTGGTTCGGCGTGTTACATGG   960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1944 GACGGTGCTGCTTCTTAACTAACACACGCGGCCCGCCCGCCAAGGCAACTGGTTTGGCCGTGTTACATGG 2003

Qy        961 ATGAATAGGTCACCGGGGTTCACCGAGACAGTGGGGGGGCCCCCCGGGGCGCCCGTAGCG   1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2004 ATGAATAGGCACTCTGGGTTCACCAAGACAGTGGTCGCCACGTGTGGGGGGGGTTC           2063

Qy       1021 GGCAACAACACTTTGATCTGCCCCACGGAGTGCTTCCGGAAGCATCCGACGTGTACATGG   1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2064 GGCAATAAAACCTTGACCTGCCCCCACGGGAACTGCTTCCGGAAGCACCCCGAGGCCACTTAC    2123

Qy       1081 ACCAAATGCCGTTCGGGCCCTTGGTTTG   1107
              ||||||||||||||||||||||||||||
Db       2124 ACCAAGTGTGGTTCGGGCGCCCTTGGGTTG  2150
```

```
Db       1764 AATGAGTCACACAGTCTCGGACCAGAGAGGCCTTATTGTTGGCACTACGGCCACCCCGGGCCGTGC

RESULT 8
ABK91430
ID   ABK91430 standard; DNA; 9605 BP.
XX
AC   ABK91430;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 7.
XX
KW   HCV; dsj; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             342..9374
FT                   /*tag= a
FT                   /product= "HCV polyprotein"
FT                   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT                   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation        replace(6931,C)
FT                   /*tag= b
XX
PN   WO20020259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
DR   WPI; 2002-599793/64.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
```

```
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cell or which containing a functional HCV replicon
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and host cell interactions, producing HCV RNA and HCV
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1910 A; 2882 C; 2733 G; 2080 T; 0 U; 0 Other;

Query Match           83.8%;  Score 927.8;  DB 6;  Length 9605;
Best Local Similarity  89.9%;  Pred. No. 5.1e-256;
Matches  995;  Conservative  0;  Mismatches  112;  Indels  0;  Gaps  0;

Qy      1  TCCTCCCGCTGCGTGCCGCGGCCCCATTGACCACGTTCGATCAGGAGGTGGGGGTCGCTCCATCCACTTAT   60
           ||||| |||| || ||||  || || ||  | ||||||| |||||||| ||||||| || ||||||||| |
Db   1044  TCCTTCCCGGTGCTGGGTACGGCTCACTCCCAGCGTTCGCGGTGGGAACCGTAGCGTCCC          1103

Qy     61  ACTGCGGACAAATAGCGAACGCCACGTGCGATTTGCTCGTTCGTTCGGTTCGCTCCGGCT          120
           || | |||| || ||| || |  ||||||  |||  |||||  |||   |||  || |||
Db   1104  ACTACGACGGATACGACGCCATGTCGATTTGCTTGTTCGTTCGGGCGCTGCTCTCTGCGGCT          1163

Qy    121  ATGTACGTGTGGGGGGATCTCTGCGGGATCTGTTTTTTCCTCGTGTTCACCTTCTCG          180
           || |||||  |||||||||||  |||||||| | ||  |||| |||||||||||||||
Db   1164  AGTACGTGGGAGAATCTGGGGATCTGTTTTCCTCGTGCCCCAGCTGTTCACTTCTCG          1223

Qy    181  CCTCGCCGACATCAGACACGGTACAGGACTCCAATTGTTCAAATCTATCCCGGCCACGTATCA          240
           || ||| || || |||||||| ||| || | |  ||| || |||| || || || ||||
Db   1224  CCTCGCCGCGACACGAGACAGTACAGGGACTGCAATTGCTCAATATATCCCGGCCACGTGACA          1283

Qy    241  GGTCACCCGCCATGGCTTGGGAATGATGATGAATGGGTCACCTACCAGCCACCCTAGTGGTA          300
           ||||||||  |||||||| ||||| || || || |||| ||||||||| ||||| ||||||
Db   1284  GGTCACCGTATGGCTTGGGATAGCATGGTCACCTACCAGCCACCCTAGTGGTA          1343

Qy    301  TCGGCAGCTACTCCCGGATCCCACAAGCTGTCGTGTGGACATGCGGGGGGCCCACTGGGGGA          360
           | ||||| ||| |||||||||||||||  |  |||| || ||| || || |||| ||||||
Db   1344  TCGGCAGTTACTCCCGGATCCCACAAGCGTGTCGTCGGTGGATATGGTGGGGCCCCATTGGGGA          1403

Qy    361  GTCCTGGCGGGCCTTGCCTTACTATTCAATGGTGGGGAACTGGCCTAAGGTCTTGGGTGTG          420
           ||||  || || ||||||  |||| |||| ||||| ||| || |||| |||||  ||||||
Db   1404  GTCCTAGCCGGCCCTTGCCTTACTATTCCATGGTGGGGGAACTGGGCTTCGATTGTG          1463

Qy    421  ATGCTACTTTTGCCGGCGCGTTGACGGGGGAACCTTACCGACAGCGGGGGACCACGGGCCGC          480
           |||||||||||||||||  |||| || |||| |||| ||| ||||| || |||| |||||
Db   1464  ATGCTACTCTTTGCCGGCGGTTGAGGGGGAACCTTATGTACACAGGGGGACGGATGCCGGGCAAA          1523

Qy    481  GCCCCCCAGGGGCTTACACCCTCTTCACACCTGGGCCGCCTCAGAAAAATCCAGCTTGTA          540
           || ||||||||||||||| || |   |||  ||||  ||||||||||||||||||| | |
Db   1524  AACACCCTCGGGATTACGTCCCTCTTTTCACCCGGGGTCATCCCAGAAAAATCCAGCTTGTA          1583

Qy    541  AACACCAACGGACAGCTGCCACATCACAACAGAAACTGCCTTGAAACTGCAATGACTCCCTCCAA          600
           ||||||| |||  |||||| ||||| ||| ||  || ||| || ||||||||||||| ||||||
Db   1584  AACACCAACGGACAGCTGGCACATCAACAGACTGCCCTGAAACTGCAATGACTCCCTCAC          1643

Qy    601  ACTGGGGGTTCCTTGCCGGCGCTGTTGTTCTACCGGCACCCAAGTTCAAACTACATCTCGGATGGCTCAGAG          660
           || ||||| ||| |||||||| |||||  |||  ||||||||||||||||||| |||| |||| |||||||
Db   1644  ACTGGGTTCCTTCCTGGCTGCGCCGGGTGTTCTACGGCACCAGGTTCAACTACTCGATGCCCAGG          1703

Qy    661  CGCATGGCCAGCTGCCGGCCCCCATTGACCACGTTCGATCAGGAGGTGGGGGTCGCTCCATCCACTTAT          720
           |||||||||||| | ||| |||||||| |||||||| |||||||| ||||||| || ||||||||| |
Db   1704  CGCATGGCCCAGCTGCCGAGCGCCCCATCGACGCTTCGCGGTGGGAACCGTAGCGTCCC          1763

Qy    721  AATGAGTCCACGCGGCTTGGACCAGAGGGCCCCTATTCGTGGGGACTACCGGCACCCGCGTGT          780
           |||| | || || ||||| ||||||||||||| |||||||||||||| | ||  || |||
Db   1764  AATGAGTCACACGGCTCGGGACCAGAGGGCCCTTATTCGTGGGGCACTACGGCACCCGCGTTACGGGT          1823

Qy    781  GGTACGTGCCCGCGTTGCAGGTGGTGTCGGGGCCACGGTAGCTGTTTCACTCCAAGCCCTTGTT          840
           |||| |||  ||| || ||||||||| |||| ||| ||| ||| || |  |  ||||||| ||
Db   1824  GGTATCGTACCCGGCGGCAGGTGTCGGCCACGGGTGTTCGGGTCCAGGTGTACGTCTTCACCCGAAGCCTTAC          1883

Qy    841  GTGGGTGGGGGACGACCGGATCGTGTTCGGGGCCCCCACGTACGTACATGAGAAGTGAGACG          900
           |||| ||||||||||| |||||||| | || ||| ||||||||||| ||| ||| |||||||
Db   1884  GTGGGTGGGGGACGACCGGTGTGCCCCTACGGTAGCCCTACAGTTTGGGGGGAGAATGAGACG          1943

Qy    901  GACGTGCTGCTTCTCAACAACACGCGCCCGCCACGGGCCCAACGGGGGCAAACTGGTTCGGCGTGTTGGG          960
           ||||| || || | ||| || ||||||||| || || |||| ||||| ||| |||| || ||| ||
Db   1944  GACGTGCTGCTTCTTAACAACACGCGCGCGGGCCCCAAGGCCAACTGGTTCGGTTGTACGTGTTGGCATGG          2003

Qy    961  ATGAATAGCAACCGGGGTTCACCAGAGCGGTGTGGGGGGCCCCCCGTTGCAACATCGGGGTC          1020
           ||||||| || ||||| |||||| ||| |||| |  |||||| | || ||  || ||||||
Db   2004  ATGAATAGCACTGGGGTTCACCGAGACGTGGGGGGCCCCCCCGTGTAACATCGGGGATC          2063

Qy   1021  GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGAGCATCCCAGGCCCACTTAC          1080
           ||||||||||||||||||| || || || |||||||||||||| |||  || |||||||
Db   2064  GGCAATAAAACCTTGACCTGCCCCACGGACTGCTTCCGGAAGCACCCGGAGGCCACTTAC          2123

Qy   1081  ACCAAATGCGGTTCGGGCGCCTTGGTTG          1107
           |||||  ||||| ||||| ||||||||
Db   2124  ACCAAGTGTGGTTCGGGGGGCCTTGGGTTG          2150


RESULT 9
ABK91428
ID   ABK91428 standard; DNA; 9605 BP.
XX
AC   ABK91428;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 5.
XX
KW   HCV; dsj; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             342..9374
FT                   /*tag= a
FT                   /product= "HCV polyprotein"
FT                   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT                   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation        replace(6463,A)
FT                   /*tag= b
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
```

XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1909 A; 2884 C; 2733 G; 2079 T; 0 U; 0 Other;

Query Match   83.8%;   Score 927.8;   DB 6;   Length 9605;
Best Local Similarity   89.9%;
Matches   995;   Conservative   0;   Mismatches   112;   Indels   0;   Gaps   0;

Qy   1 TCCTCCCGCTGCTTGGGTACGGTACCGCTCCACTCCCACGCTCCCAAGGACGCCAGCATCCCC   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1044 TCCTCCCCGCTGCTCGGGTACGCTCCACTCCCACGCTCCCAGGAACCCTAGCGTCCC 1103

Qy   61 ACTCGGACAAATACGACGCCACGTCGCAATTTGCTCGTTGGGGGCGTTGCGCTTCTGCTTCGGCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1104 ACTACGACGAGATACGCACGGCCATGTCGGATTTGCTCGTTGGGGGCGCTTCTGCTCCGGCT 1163

Qy   121 ATGTRACGTGGGGGGGATCTCTGCGGATCGTGTTTTCCTGCTCGTCTCTCAGCTGTTCACCTTCTCG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1164 ATGTRACGTGGGGAGATCTCTGCGGATCTGTTTTCCTCGTCCGCCAGCTGTTCACCTTCCG 1223

Qy   181 CCTCGGCGACATCAGACGGTTACACGGGACTCGCAATTGTTCAATCTCATCCGCCACGTATCA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1224 CCTCGCCGGCGACGAGACAGTACAGGACTCGCAATTGCTCAATATATCCCGGCCACGTGACA 1283

Qy   241 GGTCACGGCCATGGCTTGGGGAATAGAATGATGAATGAACTGGTCTACCTACAGCCACGTGGGTA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1284 GGTCACCGTATGGCTGGGGAATAGTGAATGAAACTGGTCACCTACACAGCCCTAGTGGGTA 1343

Qy   301 TCGCGAGCTACTCCGGACATCCACAACGTGTCGTGGACATGGTGGCGGCGGCCACTGGGGGA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1344 TCGGAGTTACTCCGGACATCCCACAAGCGTGTCGTGGGATATGGTGGCGGCGCCATTGGGGA 1403

Qy   361 GTCCTGGGCGGCGGGGCCTTGCCCTACTATTCCATGGTGGGGGAACTGGGCAGGTTCTTCGGTTGTTG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1404 GTCCTAGCGGCGGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTTAAGGTTCTGATTGTG 1463

Qy   421 ATGCTACTCTTTGCCGCGGCGTTGACGGGAACCTTACACGACGAGGGGGGACACACGGCGCGC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1464 ATGCTACTCTTTGCGCGGCCGTTGACGGGGGGGAACCTTATGTGCAACAGGGAGATGGCCAA 1523

Qy   481 GCCGCCACGGGGGCTTACACTCCCTCCTTCTTCACACCTGGGGGGGGTTACACTGGGTA 540
             |||||||||||||||||||||||||||||||||||||||
Db   1524 AACACCTCGGGATTACGTCCCCTCTTTCACCCGGGGTCATCCCAGAAAAATCCAGCTTGTA 1583

Qy   541 AACACCACGGCAGCTGGGACATCAACAGGAACTGCCTTGAACTGGAACTCCCTCCAA 600
             |||||||||||||||||||||||||||||||||||||||||||||||
Db   1584 AACACCAACGGCAGCTGGGACATCAACAGGAACTGCCCTGAACTGGAACTCCCTCAAC 1643

Qy   601 ACTCGGGTTCCTTGCCGCGGTGTTCTACACGCACACAGGACGTTCAAATGCCATCACTTAT 660
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1644 ACTGGGTTCCTTGCCGCGGTGTTCTACGTGCACACAGGACGTTCAAACTGCCATCACTTAT 1703

Qy   661 CGCACATGGGCCAGCTGCGCCCCATTGACCAGTTCGATCACGGGTGGGTTCCCATCACTTAC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1704 CGCACATGGGCCAGCTGCGCCCCATCGACCAGCCGTTCGCTCAGGGGTGGGGCCCATCACTTAC 1763

Qy   721 AATGAGTCCCACGGGCCTTGGACCAGAGCCCCTTATTGGCCACCTCAACCGTGT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1764 AATGAGTCACACAGCTCCGGACCAGAGGGCCTTATTGTTGGCCACTCACGACACCGGTGC 1823

Qy   781 GGTACGTGCCCGCGTTGCAGGGTGTGTGGGCCCCAGTGACTGTTCACTCACCAGCCCTGTT 840
             ||||||||||||||||||||||||||||||||||||||||
Db   1824 GGTATCGGTACCCGCGGGCGCAGGGGGTGGTGGCCAGTGTACTGCTCACCCCAAGCCCTGTC 1883

Qy   841 GTGGTCGGACGACGACGGATCGTTTCGGGGCGCCCCCTTACGTACAGAGTAGAAGGAAGATG 900
             |||||||||||||||||||||||||||||||||||||||||||||||
Db   1884 GTGGTGGGGACGACGACCGGTTCGGGCGTCCTTACGTACGTACAGAGTACTCACCCAAGCACCTTGG 1943

Qy   901 GACGTGCTGTGCTTCTCAACAACACGGGGGCGCCCCAGGGGCGAACCTGGTTCGGTTACATGG 960
             |||||||||||||||||||||||||||||||||||||||||||||||
Db   1944 GACGTGCTGCTTCTTAACAACACGGGGGCGCCCCAGGGGCGAACTGGTTCGGTTACATGG 2003

Qy   961 ATGAGAATAGCACCGGGGTTCACCAAGACGTGTGGGGACGTGCAACATCGGGGTGGTA 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2004 ATGAGAATAGCACTGGGGTTCACCAAGACGTGGGGGGGCCCCGTGTAACATCGGGGATC 2063

Qy   1021 GGCAACAACACTTTGATCTGCCCCACGGGACTGCTTCCGGGAAGCATCCCCAGGGCCACTTAC 1080
             |||||||||||||||||||||||||||||||||||||||||
Db   2064 GGCAATAAAAACCTTGACCTGCCCCACGGGACTGCTTCCCGGAAGCACACCCGGGGCCACTTAC 2123

Qy   1081 ACCAAATGCGGTTCGGGGGCCTTGGTTG 1107
             ||||||||||||||||||||||||||||||
Db   2124 ACCAAGTGTGGTTCGGGGGCCTTGGTTG 2150

RESULT 10
ABK91425
ID   ABK91425 standard; DNA; 9605 BP.
XX
AC   ABK91425;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 2.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   342..9374
FT   /tag= a
FT   /product= "HCV polyprotein"
FT   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation   replace(3946,A)
FT   /tag= b
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-BP000526.
XX

```
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1909 A; 2883 C; 2734 G; 2079 T; 0 U; 0 Other;

Query Match             83.8%;  Score 927.8;  DB 6;  Length 9605;
Best Local Similarity   89.9%;
Matches  995;  Conservative   0;  Mismatches 112;  Indels   0;  Gaps   0;
```

```
Qy       1 TCCTCCCCGCTGCTGGGTACGGCTCACTCCCACGGCGCCAAGGACGCCAGCATCCC   60
           |||||||||||||||||||||||||||||||||| ||||||||||||| ||||||
Db    1044 TCCTCCCCGCTGCTGGGTACGGCTCACTCCCACGGCGCCAAGGACGCCAGCATCCC 1103

Qy      61 ACTACGACGGATACGACGGCCACGTGGATTTGCTCGTTCGTTCGGCCTTCTGCTCCCGCT  120
           ||| || |||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    1104 ACTACGACGGATACGACGGCCATGTCGATTTGCTCGTTCGTTGGGGCCTTCTGCTCCCGCT 1163

Qy     121 ATGTAACGTGGGGGGATCCTCTGCGGGGATCTCTGCGGCGATCCTGCCAATTGTCAATCTT  180
           |||||| ||||| || |||||||||||||||||||||||||||||||||||||||||||
Db    1164 ATGTACGTGGGAGAATCTCTGCGGCGATCTGTTTCCTCGTCGCCCAGCTGTTCACCTTCTCG 1223

Qy     181 CCTGCGGACATCAGACGGTACAGGACACTGCAAATTGTTCAATCTTATTCCCGCCAGCTATCA  240
           |||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db    1224 CCTGCGGCACGAGACAGTACAGGACTCCAAATTGCTCAATATATCCCGGCCCAGCTGACA 1283

Qy     241 GGTCACCCGCATGGCCTTGGGATGATGATGAACTGGTCACCTACCTACAGCGAGCCCTTAGTTGGTA  300
           ||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db    1284 GGTCACCGTATGGCTTGGGATATGATGAACTGGTCACCTACAGCGAGCCCTTAGTTGGTA 1343

Qy     301 TCGGCAGCTTACTCCGGATCCCACAAGCTCTGCTGGACATGGTGCGTGCGTGGACATGGTGGGGGA  360
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1344 TCGGAGTTACTCCGGATCCCACAAGCTGTGGATATGGTGGTGGGGGGCGCCCCCATTGGGGA 1403

Qy     361 GTCCTGCGGGCCTTGCCTCTATTCCATGGTGGGGACTGGGGCTTAAGGTCTTGGTTGTG  420
           ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    1404 GTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGGTTCTGATTGTG 1463

Qy     421 ATGCTACTCTTTGCCGGGGTTGACGGGGGAACCTTACACGACAGGGGGACACACGGGGCCC  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1464 ATGCTACTCTTTGCCGGGGTTACATCCCTGTTTACATCCCGGTTCAGGGGGACCTTCACTTAT 1523
```

```
DT   15-NOV-2002  (first entry)

Db    1464 ATGCTACTCTTTGCCGGGGTTACATCCCTGTTTACATCCCGGTTCAGGGGGACCTTCACTTAT 1523
           ||||||||||||||| ||| ||||||||||| || ||||||||| ||| ||||||| |||||
Db    1524 AACACCCTCGGGATTACGTACGTCCCTCTTTTCACCGGGGGTCATCCCAGAAAATCCAGCTTGTA 1583

Db    1584 AACACCAACGGCAGCTTGGGCACATCAACAGAGACTTGCCCTGAACTGCAATGACTCCCTCAAC 1643

Db    1644 ACTGGGGTTCCTTGGTTCCTTCTACGTTGCCACACGGCACCAGGTTCAAACTCAGAG 1703

Db    1704 CGCCATGGCCAGCTGCAGCCCCATCCAGGCGTTGCGTCAGGGTGGGGGCCCCATCACTTAC 1763

Db    1764 AATGAGTCACACAGCTCGGAACCCAGAGGCCCTTATTCTGTTGGCACTACGGCCACCCGCTGC 1823
```

XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV
PT   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the Indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1911 A; 2883 C; 2732 G; 2079 T; 0 U; 0 Other;

Query Match   83.8%; Score 927.8; DB 6; Length 9605;
Best Local Similarity   89.9%; Pred. No. 5.1e-256;
Matches 995; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy   1 TCCTCCCGCTGCTGGGTAGCGCTCACTCCCAACGGACGCCAAGGATCCCGCT 60
              ||||||||||||||||||||||||||||||||||||||| ||||| ||||||
Db   1044 TCCTCCCGCTGCTGGGTAGCGCTCACTCCCAACGGACGCCAAGGATCCCC 1103

Qy   61 ACTCGGACAAATACGCGCACGGTACAGACTCGCAATTTGCTACTTGTTTCTCG 120
              ||||||||| |||||||| | ||||||||||||| |||||| ||||||||||
Db   1104 ACTACGACGATACGACGCCAATGTCGATTTGCTCGTTGGGGCGGGGCCTTCACCTTCTCG 1163

Qy   121 ATGTACGTGGGGGGATCTCTGCGGATCTGTTTTTCCTCGTCGTTCACTTCTCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1164 ATGTACGTGGGGGGATCTCTGCGGATCTGTTTTTCCTCGTCGTTCACTTCTCG 1223

Qy   181 CCTCTGCGCGACATCAGACGGTACAGAGACTCGAATTGTTCAATATCTACCGGCGTATCA 240
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1224 CCTCCGCCCGGCACGAGACAGTACAGGACTGCAATTGCTCAATATATCCCGGCACGTGACA 1283

Qy   241 GGTCACCGGCCATGGCTTGGGATATGATGAGGTGAAACTGGTCACCTACAGCAGCCCTAGTGGTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1284 GGTCACCGTATGGCTTGGGATATGATGAGGTGAAACTGGTCACCTACAGCAGCCCTAGTGGTTA 1343

Qy   301 TCGCAGCTACTCCGGATCCCACAAGCTGTCCTGGACATGGTGGGGGGGGGCCCACTGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1344 TCGCAGCTTACTCCGGATCCCACAAGCTGTCGTGGATAATGGTGGCGGGGGCCCATTGGGGTG 1403

Db   1344 TCGCAGCTTACTCCGGATCCCACAAGCTGTCGTGGATAATGGTGGCGGGGGCCCATTGGGGTG 1403
              |||||| ||||||| | || ||||||||||||||||||||||||||||||||
Db   1404 GTCCTTAGCGGCGCCCTTGCCTTACTATTCCATGGTGGGGGAACTGGGCTTAAGGTTCTGATTGTG 1463
Db   1464 ATGCTTACTCTTTTGCCGCGGCGTTGACGGGGGAACCTTACGACGAGGGGGGACACAGCCCGC 1523
Db   1524 GCCGCCACCGGGCGTTACATCCCTCTTCACACCTGGGCCGGGCTTCAGAAAATCCAGCTTGTA 1583
Db   1584 AACACCCTCGGGATTACGTCCCTCTTTTCACGGGGTCATCCCAGAAAATCCAGCTTGTA 1643
Db   1644 ACTGGGGTTCCTTGCCTGCGCGTCTCTTTCACGGTGCACAAGAGTTCAAGTCCATCTGGATGCCCAGAG 1703

Qy   361 GTCCTTAGCGGGGGGCCCTTGGCCTTCACTATTCCATGGTGGGGGGAACTGGGCTTAAGGTTCTGATTGTG 420
              |||||| ||||||| | || |||||||||||||||||||||||||||||||||||||
Db   1704 CGCATGGGCCAGCTGCCGCCCCATTGACCAGTTCGATCAGCAGGGGGTGGGGTCCCATCAC 1763

Qy   421 ATGCTACTCTTTTGCCGCGGCGTTGACGGGGGAACCTTACGACGAGGGGGGACACAGCCCGC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1764 AATGAGTCACACAGTCTCGGACCAGAGGGCCTTATTGTTGGCACTAGCACCCCGGCGTGC 1823

Qy   481 GCCGCCACCGGGCGTTACATCCCTCTTCACACCTGGGCCGGGCTTCAGAAAATCCAGCTTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1824 GGTATCGTACCCGCGGTGTTGCAGGTGGTGGGCCAGTGATCGTCAGTGAATTGGTAAAATGAGAC 1883

Qy   541 AACACCCAACGGACGCAGCATGGGCGCATCAACACAGAACTGCCTTGGAACTGCAATTGACTCTCCTTCCA 600
              ||||||| |||||||||||||||||||||| |||||||||| |||||||||||||||||||||||
Db   1884 GTGGGTGGGGACGACCGGTTCGGCGGTCCCTAGGTACACAGTTGGGGGGAGGAATAGAGACG 1943

Qy   601 ACTGGGGTTCCTTGCCTGCGCGTCTCTTTCACGGTGCACAAGAGTTCAAGTCCATCTGGATGCTCAGAG 660
              ||||||| |||||| |||||| |||||||||||||||||||||||||||||||| |||||||
Db   1944 GACGGTCGTGCTTCTTAACACACGCGCCGCCGGGAACTGGTTTGGCGTTGACATGG 2003

Qy   661 CGCATGGGCCAGCTGCCGCCCCATTGACCAGTTCGATCAGCAGGGGGTGGGGTCCCATCAT 720
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||| ||||||
Db   2004 ATGAATAGCACCGGGTTCACCAGAGGTGTGGGGGCCGTGCAAGATGCCCGGTCG 2063

Qy   721 AATGAGTCCACAGCGGCTTGGACCAGAGGGCCCTTATTGCTGTCGGACTACGCACTTCAACGGTGT 780
              |||||||| |||||||||||||||||||||||| |||||||||||||||||||| ||||||
Db   2064 GGCCAATAAAAACCTTGACCTGCCCCACGGGACTGCTTCCGGGGACTTACCCCGGAAGCCACTTAC 2123

Qy   781 GGTATCGTGCCCGCGGTTGCAGGTGGTGGGCCAGTGTACTGTGTACTACTGGTACTCTCCAAGCCCTGTT 840
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||||||
Db   2124 ACCAAGTGTGGTTCGGGGCCTTGGTTG 2150

Qy   841 GTGGGTGGGGACGACCGGATCGTTTCGGCGCCCCTACGTACACAGATGGGGTGAGAATGGACG 900
              ||||||||||||||||||||||||||||||||| ||||||||| |||||| |||||||||||
Db   1884 GTGGGTGGGGACGACCGGTTCGGCGGTCCCTAGGTACACAGTTGGGGGGAGGAATAGAGACG 1943

Qy   901 GACGTCGTGCTTCTCTACAACACGGCGCCGCCGCCACGGGGGCGCAACGGTGGTTGGCGTTGACATGG 960
              |||| |||||||| |||||||| |||||||||| | |||||||||||||| ||| |||||||||
Db   1944 GACGGTCGTGCTTCTTAACACACGCGCCGCCGGGAACTGGTTTGGCGTTGACATGG 2003

Qy   961 ATGAATAGCACCGGGTTCACCAAGAGGTGCAAACATGGGGGGGTC 1020
              ||||||||||||||||||||||||||||||| ||||||| |||||
Db   2004 ATGAATAGCACCGGGTTCACCAGAGGTGTGGGGGCCGTGCAAGATGCCCGGTCG 2063

Qy   1021 GGCCAACAACACTTTGATCTGCCCCACGGACGACTGCTTCCGGGGAGCACATCCCGAGGGGCCCACTTAC 1080
              |||||||| |||||| ||||||||||||| |||||||||||||||| |||| ||||||||||||||||||
Db   2064 GGCCAATAAAAACCTTGACCTGCCCCACGGGACTGCTTCCGGGGACTTACCCCGGAAGCCACTTAC 2123

Qy   1081 ACCAAAATGCGGTTCGGGGGCCTTGGTTG 1107
              |||||||||| |||||||||||||||||||
Db   2124 ACCAAGTGTGGTTCGGGGGCCTTGGTTG 2150

RESULT 12
ABK91433
AC   ABK91433
XX
ID   ABK91433 standard; DNA; 9605 BP.
XX
AC   ABK91433;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 10.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers

```
FT   CDS              342..9374
FT                    /*tag= a
FT                    /product= "HCV polyprotein"
FT                    /note= "The polyprotein consists of the Core, E1, E2, P7,
FT                    NS2, NS3, NS4A, NS4B, NS5A, NS5A and NS5B proteins"
FT   mutation         replace(6953,C)
FT                    /*tag= b
XX
PN   WO20025093Z1-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9605 BP; 1910 A; 2882 C; 2733 G; 2079 T; 0 U; 1 Other;

Query Match             83.8%;  Score 927.8;  DB 6;  Length 9605;
Best Local Similarity   89.9%;  Pred. No. 5.1e-256;
Matches 995;  Conservative  0;  Mismatches 112;  Indels  0;  Gaps  0;

Qy           1 TCCTCCCGGCTGGTAGGCGCTCACTCCCAGTCCAGCGACGGCCAGGGCCGCT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1044 TCCTCCCGGCTGCTGGGTAGGCGCTCACTCCCAGTCCAGCGACGGCCTAGGGCGGCTCCC 1103

Qy          61 ACTCCGACAAATACGACGCCACGTCGCAATTTGCTCGTTGGGGCGGCGGGCCCTT 120
               |||| ||| | |||| ||||||||||||||||||||||||||||||||||||
Db        1104 ACTACGACCGGATACGACGCACCATGTCGATTTGCTCGTTGGGGCGGCGGGTCTGCTCGTTGCGGT 1163

Qy         121 ATGTACGTGGGGGGGATTCTCTGCGGAATCTGTTTTCTCGTCTCTCAGCGTTCACCTTCTCG 180
               |||||||||||||||||| |||| ||||||||||||||||||||||||||||
Db        1164 ATGTACGTGGGAGATCTCTGCGGATCTGTTTTCCCGTCGCCCAGCTGTTCACCTTCTCG 1223

Qy         181 CCTCGCCGCCGCCACGAGACAGTACAGGACTGCAATTGTTCAATCTATCCCGGGCCGCCACGTTCA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1224 CCTCGCCGCCGCCACGAGACAGTACAGGACTGCAATTGCTCAATAATATCCCGGGCCCGCCACGTGACA 1283
```

```
Qy         241 GGTCACCGCATGGCCTTGGGGATAATGATGAATCTGGTCACTTACCTTACTCAGCATCCAGCCCTAGTGGTA 300
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1284 GGTCACCGTATGGCTTGGGGATATGATGAATCTGGTCACTTACCTTACTCAGCCCTAGTGGTA 1343

Qy         301 TCGCAGCTACTTCCGGATCCCACAGCGTCGTCGTTGGACGGCATTGGTGGGGCCCCACTGGGGA 360
               |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        1344 TCGCAGTTACTCCGGATCCCACAGCGTCGTCGTTGGACGGCATTGGTGGGGCCCCATTGGGA 1403

Qy         361 GTCCTGGCGCGGCGCCTTGCCTACTATTCCATGGTGGGGAAACTGGTCTTGGTTGTG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1404 GTCCTTAGCCGGCGCCTTGCCTACTATTCCATGGTGGGGAAACTGGTTCTGATTGTG 1463

Qy         421 ATGCTACTCTTTGCCCGGGCGTTGACGGGGGAACCTTACGACGACACACGGCCCGC 480
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        1464 ATGCTACTCTTTGCCCGGGCGTTGACGGGGGAACCTTACGACGACACACGGCCCAAA 1523

Qy         481 GCCGCCCACGGGGCTTACACTCCCTCTTCACCACCTGGGGCGGCGGGACGGCCAA 540
               || | |||||||||||||||||||||||||||||||||||||||||||||||
Db        1524 AACACCCTCGGGATTACGTCCCTCTTTACCACCTGGGGTCAGAAATCCAGCTGTA 1583

Qy         541 AACACCCAACGCGCAGTGTGGCACATCAACACAGAACTTGCCTTGAACTTGCAAATGACTTCCCCA 600
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1584 AACACCCAACGCGCAGCTGGCACATCAACACAGGACTGCCCCTGAACTGCAAATGACTTCCCTCAAC 1643

Qy         601 ACTCGGGGTTCCTTGCCCGGGCGTTCTACACGGCACAGGTTCAATGGCTCAGAG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1644 ACTCGGGGTTCCTTGCTCGCGCGTTCTACACGGGCACAAGTTCAACTCATCTGAGTCGGATGCCCAGAG 1703

Qy         661 CGCATGGCCAGCTGCCGCCCCCATTGACCAGTTCGATCAGGGGGTCCATCACTTAT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1704 CGCATGGCCAGCTGACCAGTTCGATCAGGGGTTTGGCCTGCCCTCCTCAACCACTTAC 1763

Qy         721 AATGAGTCCCACGCGTTGGACCAGAGGCCCTTATTCTGCTGGCACTACCACGTGT 780
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1764 AATGAGTCACACAGCTCCGACCAGAGGCCCTTATTGTTGGCACTACCGACCCCGTGC 1823

Qy         781 GGTATCGTGGTGCCCGGGTGCAGGTGTGGCCCAGTGTACTGTTTCACTCCAAGCCCTGTT 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1824 GGTATCGTACCCCGCGCAGGTGTGTGCCAGGGTACTGCTTCACCCCAAGCCCTGTC 1883

Qy         841 GTGGTGGGGACGGACCGATCGTTTCGGCGCGCCCCCTACGTACAGATGGGGTGAGAGAC 900
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1884 GTGGGTGGGGACGGACCGACCGGCTTCCCGGGCTCCCTACGTTACAGTTGGGGAGAAATGAGAGAC 1943

Qy         901 GACGTGCTGCTTCTCAACAACACGCGGGGCCCCCCACGGGGCAACTGGTTTGTTCGGTTCATGG 960
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1944 GACGTGCTGCTTCTTAACACACGCGCGGGGCCCCGCAAGGCAACTGGTTTGCTGTGTACATGG 2003

Qy         961 ATGAATAGAGCACCGGGTTCACCAAGAGTGTGGGGGGCCCCGGGGGGGCGGTC 1020
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2004 ATGAATAGGCACTGGGTTCACCAAGAGACGTGGGGGGCCCCGTGGTAACATCGGGGGGGGATC 2063

Qy        1021 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGGAAGCGAGGCCCACTTAC 1080
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2064 GGCAATAAAACCTTGACCTGCCCCAGGGACTGCTTCCGGGAAGCGCACCCCGAGGCCCACTTAC 2123

Qy        1081 ACCAAAATGCGGTTTCGGGGGCCCTTGGTTG 1107
               |||||||||||||||||||||||||||||||
Db        2124 ACCAAGTGTGTTCGGGGGCCCTTGGTTG 2150
```

```
RESULT 13
AAD25332

ID   AAD25332 standard; cDNA; 9605 BP.
XX
AC   AAD25332;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Hepatitis C virus (HCV) full-length cDNA mutant #2.
XX
KW   Hepatitis C virus; HCV; transfection; infection; virus neutralisation;
```

KW   gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;
KW   mutant; ss.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200189364-A2.
XX
PD   29-NOV-2001.
XX
PF   23-NOV-2001; 2001WO-US016822.
XX
PR   23-MAY-2000; 2000US-00576989.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Rice CM, Blight KJ;
XX
DR   WPI; 2002-066755/09.
XX
XX
PS   Example 2; Page 97-100; 174pp; English.
XX
PT   Hepatitis C virus variants having greater transfection efficiency and
PT   ability to survive subpassage, useful as a vaccine for immunizing primate
PT   to the virus, comprise non-naturally occuring viral sequences.
XX
CC   The invention relates to Hepatitis C virus (HCV) variants which include
CC   polynucleotides comprising non-naturally occurring HCV sequence and HCV
CC   variants that have a transfection efficiency and ability to survive
CC   subpassage greater than HCV that have wild-type polyprotein coding
CC   regions. The polynucleotides of the invention are useful for identifying
CC   a cell line that is permissive for infection with HCV and detecting
CC   replication of HCV in cells of the cell line. They are also useful for
CC   testing a compound for anti-viral properties and for inhibiting HCV
CC   infection. They are also useful for the generation of defined HCV virus
CC   stocks to develop in vitro and in vivo assays for virus neutralisation,
CC   attachment, penetration and entry, structure/function studies on HCV
CC   proteins and RNA elements and identification of new antiviral targets, a
CC   systematic survey of cell culture systems and conditions to identify
CC   those that support wild-type and variant HCV RNA replication and particle
CC   release, production of adaptive HCV variants capable of more efficiency
CC   replication in cell culture, production of HCV variants with altered
CC   tissue or species tropism, establishment of alternative animal models for
CC   inhibitor evaluation including those supporting HCV variant replication,
CC   development of cell-free HCV replication assays, production of
CC   immunogenic HCV particles for vaccination, engineering of attenuated HCV
CC   derivatives as possible vaccine candidates, engineering of attenuated or
CC   defective HCV derivatives for expression of heterologous gene products
CC   for gene therapy and vaccine applications and for utilisation of the HCV
CC   glycoproteins for targeted delivery of therapeutic agents to the liver
CC   or other cell types with appropriate receptors. Vaccine comprising these
CC   sequences is useful for inducing immunoprotection to HCV in a primate.
CC   The present sequence is Hepatitis C virus (HCV) full-length cDNA
CC   containing the mutation that results in Ser to Ile at position 1179 of
CC   HCVrep1bartman protein
XX
SQ   Sequence 9605 BP; 1910 A; 2884 C; 2731 G; 2080 T; 0 U; 0 Other;

Query Match   83.8%; Score 927.8; DB 6; Length 9605;
Best Local Similarity   89.9%; Pred. No. 5.1e-256;
Matches 995; Conservative   0; Mismatches 112; Indels 0; Gaps 0;

Qy   1 TCCTCCCGCTGCTGGGTAGCGTCACTCCCACGCCGCTCGCGGGCCAAGGAGACGGCCAGCAATCCCC 60
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||
Db   1044 TCCTCCCGGCTGCTGGGTAGCGCTCACTCCCACGCTTCGATTTCGTTCGTTGGGGTGTCGCTGTCCGCT 1103

Qy   61 ACTCGGACAAATACGACGGCCACGTCACTTCCTCGTTCGGTTGGGCGGGTGCCTTTCTGCCTGGGCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1104 ACTACGACGATACGACGCCATGTCGATTTGCTCGTTGGGGGCGGGGCCGCGAAGCGTAGCGTCCCC 1163

Qy   121 ATGTACGTGGGGGGATCTCTCGGAGATCTTGTTTCCTTGTTTCCTTCCCGTCGGTTCACCTTCTCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1164 ATGTACGTGGGGAGAGATCTCGGCGATCGTTTCTCCGTGGTCGTTCGCCCGTCGGTTCACCTTCTCTG 1223

Qy   181 CCTCGCCCGACATCAGACGGTACAGGACTGCAATTGTTCAATTCTTACCCCGGCCACGTATCA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1224 CCTTCGCCCGGCCACGAGACAGTACAGGACTGCAAATTGCTCAATTATTCCCCGGCCACGTGACA 1283

Qy   241 GGTCACCGCCATGGCTTGGGATATGGATGAGAACTGGTCAACCTACCTAGCAGCGACGCCTAGTGGTA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1284 GGTCACCGTATGGCTTGGGATATGGATGAGAACTGGTCAACCTACCTAGCAGCGACGCCTAGTGGTA 1343

Qy   301 TCGCAGCTACTCCGGATCCCCACAAGCTGTCGTGGACATGGTGGGCGGGCGGGGGCCACTGGGGA 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1344 TCGCAGTTACTCCGGATCCCCAACAAGCTGTCGTGGTGGATATGGTTGGGCGGGCGGGGGCCATTGGGGA 1403

Qy   361 GTCCTTGGCCGGGGCCCTTGCCTTACTATTCCCATATGGTGTGGGGGAACCTGGGGGCGGGGCTAAGG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1404 GTCCTTAGCGGGGCCCTTGCCTACTATTCCCACAACTGGTGTGGGGGAACCTGGGGGGCTAAGGGTCTTGGTG 1463

Qy   421 ATGCTACTCTTTGCCGGGCCGTTGACAGGGGGAACCTTACACGACCAGGGGGGAACCTTACTGATTGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||| |||
Db   1464 ATGCTACTCTTTGCCGGGCCTTGCCTACTATTCCCATAATGTTGACACGGGGGGAACCTTACTGATTGTG 1523

Qy   481 GCCGCCCACCGGGCCGTTACATCCCTCTTTCACACCGGGGCCGCGGGGCGGTCAGAAAATCCAGCTTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1524 AACACCCTCGGGGATTAGACGTGCCCTCTTTTCACCCGGGGTCATCCCAGAGAAAATCCAGCTTGTA 1583

Qy   541 AACACCAACCGGCAGCTCGGACCATCAACACAGAACTGCCTTGAACTTGCAATGCTCCCTCCAA 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1584 AACACCAACCGGCAGCTCGGACCATCAACACAGAACTGCCCTGAACTTGCAATGACTCCCTCAAC 1643

Qy   601 ACTGGGGTTCCTTGCCCGGCGCTGTTCTACACGCACCACAGGTTCAATGCGTTCCGGATGTCCAGAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1644 ACTGGGGTTCCTTGCTGTCGCGTGTTCTACCGTGTCACAAGTTCAACTCATCTGGATGTGCCCAGAG 1703

Qy   661 CGCATGGCGCCAGCCTGCGCGCCCCCATTGACCAGTTCGATCAGCAGGGGGGGTCCATCACTTAT 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1704 CGCATGGCCGCCAGCGCTGCGAGCCCCATCGACACGGCGGTTCGGCTACCGGCCCATCACTTAC 1763

Qy   721 AATGAGTCCCACGCTTGGACCAGAGGCCCTTATGTGCTGCCACTCACCACCCGCGTGT 780
              ||||||||||||||||||||||||| ||| |||||||||||||||||||||||| |||
Db   1764 AATGAGTCACACAGTCTCGACGCCTTATTGTTGGCACTACCACCCGCGTGGC 1823

Qy   781 GGTATCGTGCCCGGCGGTTGCGAGGTGTTGTGGCCCCAGTAGCTGTTTCCACTCCAAGGTGTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1824 GGTATCGTACCCGCGCGCAGGTGTTGTGGTTCAGTGTACTGCTTCACCCCGAAGCCCTGTC 1883

Qy   841 GTGGTGGGGACGGACCGATCGTTTCCGCGCCCCTAGGTACAGATGGGGGTGAGAATGAGACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1884 GTGGTGGGGACGGACCGATCGTTTCGGCGCCCCTTACGTACACAGTTGGGGGGGAGAATGAGACG 1943

Qy   901 GACGGTGCTGCTTCTTCAACACACCGGGCGCCCACCGGGGCAACTGGTTTGGCGTGTTACATGG 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1944 GACGGTGCTGCTTCTTCAACACACCGGGCGCCCAAGGCCAACTGGTTTGGCGTGTTACATGG 2003

Qy   961 ATGAATAGCACCGGGGTTCACCAAGAGCGTGGGGGGCCCCCGTGCAACATCGGGGGGGGTC 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2004 ATGAATAAGCACTGGGGTTCACCGAGACGTGCGGGGCCCCGTGTAACATGGGGGGGGATC 2063

Qy   1021 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGAAGCATCCCGAGGCACTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2064 GGCAATAAACCTTGACCTGCCCCACGGACTGCTTCCGGAAGCACCCGAGGCACTTAC 2123

Qy   1081 ACCAAAATGCGGTTCCGGGGCCCTTGGTTG 1107
              ||||| ||||||||||||||||||||||||
Db   2124 ACCAAGTGTGTTCCGGGGCCTTGGTTG 2150

XX
AC   ABK91427
XX
ID   ABK91427 standard; DNA; 9608 BP.
XX
AC   ABK91427;
XX
DT   15-NOV-2002 (first entry)

RESULT 14
ABK91427

```
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 4.
XX
KW   HCV; dej; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             342..9374
FT                   /*tag= a
FT                   /product= "HCV polyprotein"
FT                   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT                   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation        replace(6458..6462,GA)
FT                   /*tag= b
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-597993/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more HCV
CC   NS3, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic
CC   acids; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC   mutant of the invention. Note: The present sequence is not shown in the
CC   specification but was created by the indexer using the HCV sequence
CC   appearing as ABK91411 and the information in Claim 9
XX
SQ   Sequence 9608 BP; 1913 A; 2883 C; 2733 G; 2079 T; 0 U; 0 Other;

Query Match    83.8%;  Score 927.8;  DB 6;  Length 9608;
Best Local Similarity  89.9%;
Matches  995;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;
```

```
Qy        61 ACTGCGACAATACGGACGGCCACGTCGATTTGCTCGTTGCTTCGTTGGGGGGTTGGGGGGG  120
             ||||  |||| || |||||  || ||||||||||||||| |  ||| | | ||||  ||
Db      1104 ACTACGGACGATACGGACGCCCATGTCGATTTGCTCGTTGGGGTCTCTCAGCTGCTTCACCTTCTCG 1163

Qy       121 ATGTACGGTGGGGGGGATCTTCGGGGATCTGTTGTTTTCCTGGTCGTTGTTCAGCTCTCG  180
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db      1164 ATGTACGTGGGAGATCTTCTGGGGATCTGTTGTTTTCCTGGTCGTTGTTCAGCTTGGTA 1223

Qy       181 CCTCGCCGGACAATCAGACGGTACAGGACTGCAAATTGTTCAATTATCCCGGCGTATCA  240
             |||| ||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db      1224 CCTCGCCGGCACGAGACGGTACAGGACTGCAAATTGCTCAATATATCCCGGCCGCAGTGACA 1283

Qy       241 GGTCACCGCCATGGGGGGGATATGGGGATATGAATGAAGAAGGACTGTCACCTACAGCAGC  300
             ||||||||||||||||||||||||||||||| ||||||||||||| |||||||||||||
Db      1284 GGTCACCGTATGGGTTGGGGGATATGAATGATGAACTGGTCACCTTACAGCAGCCCTAGTGGTA 1343

Qy       301 TCGGCAGCTACTTCCCGGATCCCAACAAGCTGTCGTGGATCCCAACATGCGGGGGGGGA  360
             |||| |||||||||||||| ||||||||||||||| |||||||| |||||||||||||
Db      1344 TCGCAGTTACTCCGGATCCCACACAAGCTGTCGTGGATGATGTTGGCGGGGGCCCATTGGGGA 1403

Qy       361 GTCCTGGCCGGGGGCCTTGCCTACTATTCCATGGGTGGGGGGAACTGGGGGGCTAAGGTCT  420
             |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db      1404 GTCCTTAGCCGGCGCCTTGCCTACTATTCCATGGTGGGGGGAACTGGGGGCCTAAGGTTC 1463

Qy       421 ATGCTACTTCTTTTGCCCGCGGCGGTTGACCGGGGGGGACCTTACACGACGCACACGCAC  480
             ||||||| |||||||||| ||||||| |||||||||||||||||||| |||||||||
Db      1464 ATGCTACTTCTTTTGCCCGCGGTTGACCGGGGGGGACCTTATGCACGACGTCGAA 1523

Qy       481 GCCGCCCACCGGGGCTTACATCCCTCTTCACACCTGGGCGCCGGCCTCAGGAAAATCCAGC  540
             |||| ||||||||| |||||||||||||||||| ||||||||||||| ||||||||||
Db      1524 AACACCCTCGGGATTACGTCCCTCTTTTCACCGGGTCATCCCAGAAAATCCAGCTTGGTA 1583

Qy       541 AACACCAAGCGCAGCTGGGCACATCAACAGAACTGCCTTGAACTGCAATGACTCCCTCCAA  600
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      1584 AACACCAAACGGCAGCTGCGCACATCAACAGGAACTGCCCTGAACTGCAATGACTCCCTCAAC 1643

Qy       601 ACTGGGGTTCCTTGCGTGTGTTCTACACGCACCAGGTTCAATGCGTCGCGGATTGCTGAGAG  660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1644 ACTGGGGTTCCTTGCGTTCTACGTGCACAAGTTCATCTGGAATGCCAGAG 1703

Qy       661 CGCATGGCCAGCTGCCGGCCCCATTGACCAGTTCGATCAGGGGTGGGGGGTCCCATCACTTAT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1704 CGCATGGCCAGCTGCGCCATCGACAGCGTTCGCTCAGGGGTGGGGGCCCATCACTTAC 1763

Qy       721 AATGAGTCCCACGGCTTGGACGAGGGCCCTATTGCTGCTGGCGCCACTACCGCCCTCAACCGTGT  780
             |||||||| |||||||||||||||||||| |||||| ||||||||||||||||||||
Db      1764 AATGAGTCACACAGCTCGGACGAGGGCCTTATTGTTGGGCGCACTACCGCCGCCTGC 1823

Qy       781 GGTTATCGTGCCCGCGGTTGCAGGTGTGTGGCCCCAGTGTACGTTCACTCCAAGCCCTGTT  840
             |||||||||||| ||||||||||||||||||||| ||||| ||||||| |||||||||
Db      1824 GGTATCGTTACCCGCGCGCAGGTGTGTGGTTCAGTGTACTCGTTCACCCCAAGCCCTGTC 1883

Qy       841 GTGGGTGGGGACGGACCGGATCGTTTCGGCGCCCCGCGGGCCGCCTACGTAGTGACGACG  900
             |||||||||||| |||||||||||||||| |||||||||||||||||||||||||||||
Db      1884 GTGGGTGGGGACGGACCGGATCGTTCGGGCCCCTTCGGCGTCCCTACGTAACAGATGAGACG 1943

Qy       901 GACGTCGTGCTTCTTCAACAACACGGCGCCGCCCGCCGGCAACTGGTTCGGCTGTACATGG  960
             |||||| ||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      1944 GACGTCGTTCTTCTTAACACAACGCGGGCCGCGCCGCCGCAACTGGTTTGGCTGTACATGG 2003

Qy       961 ATGAAATAAGACGACCGGGTTCACCGAAGACGTGTGGGGGGGGCCCGGGGGGGGGTC  1020
             ||||||||||| |||||||||||||| ||||||||| ||||||||||||||||||||
Db      2004 ATGAAATAGCACTGGGTTCACCGAAGACGTGCGGGGCCCGTGTAACATGGGGGGGGATC 2063

Qy      1021 GGCAACAACACTTTGATTCTGCCCCACGGACTGCTTCCGGAAGCATCCCGAGGCCACTTAC 1080
             ||||||||||||||||| |||||||||||||||| ||||| |||||||| |||||||||
Db      2064 GGCAATAAAACCTTGACCTGCCCCACGGACTGCTTCGGGAAGCCACCCCAGAGGCCACTTAC 2123

Qy      1081 ACCAAATGCGGTTCGGCGGCCTTGGTTG  1107
             |||||| |||||||| ||||||||||||
Db      2124 ACCAAGTGTGGTTCGGGGGGCCTTGGTTG 2150

Qy         1 TCCTCCCCGCTGCTGGGTAGCGCGTCACTCCCACGCGCCCAAGGACGACGCCAGCAATCCCC   60
             ||||||| ||||||||||| || |||||||||||||| |||| |||||  ||||||||||
Db      1044 TCCTCCCGGCTGCTGGGTAGGGCTCACTCCCACGCGCGCGGCGGCCCAGGAACGGTAGCCTCTCCC 1103
```

RESULT 15

AAD25331

ID   AAD25331 standard; cDNA; 11062 BP.
XX
AC   AAD25331;
XX
DT   12-MAR-2002 (first entry)
XX
DE   Hepatitis C virus (HCV) full-length cDNA mutant #1.
XX
KW   Hepatitis C virus; HCV; transfection; infection; virus neutralisation;
KW   gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;
KW   mutant; ss.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200189364-A2.
XX
PD   29-NOV-2001.
XX
PF   23-MAY-2001; 2001WO-US016822.
PR   23-MAY-2000; 2000US-0576989.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Rice CM, Blight KJ;
XX
DR   WPI; 2002-066755/09.
XX
PS   Example 2; Page 93-96; 174pp; English.
XX

PT   The invention relates to Hepatitis C virus (HCV) variants which include
PT   polynucleotides comprising non-naturally occurring HCV sequence and HCV
PT   variants that have a transfection efficiency and ability to survive
PT   subpassage greater than HCV that have wild-type polyprotein coding
PT   regions. The polynucleotides of the invention are useful for identifying
PT   a cell line that is permissive for infection with HCV and detecting
PT   replication of HCV in cells of the cell line. They are also useful for
PT   testing a compound for anti-viral properties and for inhibiting HCV
PT   infection. They are also useful for the generation of defined HCV virus
PT   stocks to develop in vitro and in vivo assays for virus neutralisation,
PT   attachment, penetration and entry, structure/function studies on HCV
PT   proteins and RNA elements and identification of new antiviral targets, a
PT   systematic survey of cell culture systems and conditions to identify
PT   those that support wild-type and variant HCV RNA replication and particle
PT   release, production of adaptive HCV variants capable of more efficient
PT   replication in cell culture, production of HCV variants with altered
PT   tissue or species tropism, establishment of alternative animal models for
PT   inhibitor evaluation including those supporting HCV variant replication,
PT   development of cell-free HCV replication assays, production of
PT   immunogenic HCV particles for vaccination, engineering of attenuated HCV
PT   derivatives as possible vaccine candidates, engineering of attenuated or
PT   defective HCV derivatives for expression of heterologous gene products
PT   for gene therapy and vaccine applications and for utilisation of the HCV
PT   glycoproteins for targetted delivery of therapeutic agents to the liver
PT   or other cell types with appropriate receptors. Vaccine comprising these
PT   sequences is useful for inducing immunoprotection to HCV in a primate.
XX
CC   The present sequence is Hepatitis C virus (HCV) full-length cDNA
CC   containing the mutation that results in Ser to Ile at position 1179 of
CC   HCVrep1bartMan protein, where the 5'NTR is fused to the neomycin
CC   phosphotransferase gene and the EMCV IRES is inserted upstream of the HCV
CC   open reading frame
XX
SQ   Sequence 11062 BP; 2217 A; 3291 C; 3145 G; 2409 T; 0 U; 0 Other;

Query Match   83.8%; Score 927.8; DB 6; Length 11062;
Best Local Similarity   89.9%; Pred. No. 5.4e-256;

Matches 995; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy   1 TCCTTCCCGCTGCTGTGGGTAGGGCTAGCGCTTCACTCCCAGCGTTCCGCGGCTCGCGGCCCAAGGGACGGCCAGCATCCCC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2501 TCCTTCCCGCTGCTGTGGGTAGGGCTAGCGCTTCACTCCCAGCGTTCCGCGGCTCGCGGCCCAAGGGACGGCCAGCATCCCC 2560

Qy   61 ACTGGGACAATACGACGACCACGTCGTGGGATTTGCTTGTCGTTGGGGGGGGCTGCCTTCGTCCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2561 ACTACGACGATACGACGACCACGTCGGATTTGCTCGTTCGTTGGGGGGGGCTGCCTTCGTCCGGCT 2620

Qy   121 ATGTACGTGGGGGATCCTCTCCGGGACTCGATTTGCTCGTCGTTCTCTCAGCTTCTCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2621 ATGTACGTGGGGGATCCTCTCCGGGAATCTGATTTGCTCGTTGGGGGGGCTCTCTCG 2680

Qy   181 CCTCGCCCGACATCAGGACGGTACAGGACTCGAAGTTCAATCTATCCCGGCCCACGTATCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2681 CCTCGCCCGGGCACGAGACAGTACAGGACTCGAAGTTCAATATATACCCGGCCCACGTGACA 2740

Qy   241 GGTCACCGCATGGCTTGGGGATATGATGAGAACTGGTCACCTACAGCAGCCCTAGTGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2741 GGTCACCGTATGGCTTGGGGATATGATGAGTGAACTGGTCACTTACAGCAGCCCTAGTGGTA 2800

Qy   301 TCGCAGCTACTCCGGATCCCACAACCTGTGTCGTTGGACATGGTCGGGGGCCCCACTCGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2801 TCGCAGCTTACTCCGGATCCCACAACCTGTCGTGGACATGGTCGGGGGCCCCATTCGGGGA 2860

Qy   361 GTCCTTGCGGGGGCCTTGCCTCACTATTCCATTGGTGTGGGGGAACTGGCGTAAGGTCTTGGTTGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2861 GTCCTTAGCGGGGCCTTGCCTACTATTCCATTGGTGGGGAACTGGGTTAAGGGTTCTGATTGTG 2920

Qy   421 ATGCTACTCTTTGCCGCGGTCCCTCTTTTTCACCGGGGTCATCCCAGGAAAATCCAGTTGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2921 ATGCTACTCTTTTGCCGCGGTTGACGGGGGGGAACCTAGTGACAGGGGGACGGATTGGCCAAA 2980

Qy   481 GCCGGCCCACGGGCCTTACATCCCTCTTTTCACACCTGGGCCGGCTCAGAAAATCCAGTTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2981 AACACCCTTCGGGATTACGTCCCTCTTTTCACCGGGGTCATCCCAGGAAAATCCAGTTGTA 3040

Qy   541 AACACCAACGGCGACGTGGGCGACATCAACAGAACTGCGCCTTGAACTGCAATGACTCCCTCCAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3041 AACACCAACGGGCACATCAACATCAACACAGACTGACTGCCCTGAACTGACTCCCTCAAC 3100

Qy   601 ACTGGGTTCCTTGCCGCGCTGGTTCTACACCGCACAGTTCAAATGCGTCCGGATGCGTCGAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3101 ACTGGGGTTCCTTGCTGCGCTGGTTCTACGTGCACAAGTTCATCTGGATGCGCCCAGAG 3160

Qy   661 CGCAATGGGCCAGGCTGCGGCGCCCCATTCAGACAGTTCGGATCAGGGGGTCGGGGGGTTATT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3161 CGCATGGGGCCAGGCTGCCAGGCCCCATTCGACCAGTTCGGGGTGGGGGGGCCATCACTTAC 3220

Qy   721 AATGAGTCCCACGGGCTTGGACGGAGGGCCCTATTGGCTGGTGGGGCACTACGCCACCACCGGTGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3221 AATGAGTCACACAGCGTCCGACCAGAGGGGCCTTATTGGTGCGACTACGCCACTCACTTAC 3280

Qy   781 GGTATCGTACCCCGCCCGCGTTTGCAGGGTGTGGGGCCCAGGGTACGTTTCAAGCCCTGTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3281 GGTATCGTACCCCGCCGCGCAGGGTGTGGGTCCAGGTGTACTTCACCCCAAGCCCTGTC 3340

Qy   841 GTGGGTGGGGACGACCGGATCGTTTCCGCGCGCCCCTACGTACAGGAATGGGTGAGGATGAGACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3341 GTGGGTGGGGACGACCGGGTCCCTACGTACAGTTGGGGCGGAGCAACTGAGACG 3400

Qy   901 GACGTGCTGCTTCTCAACAACACGCGGGCGCCCCACGGGGCAACTGGTTCGCTGACACATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3401 GACGTTGCTCGTTCTTAACACACACGCGCGCGCCCCCGCAAGGCAACTGGTTTGGCGTACATGG 3460

Qy   961 ATGAAGATAGCACCGGGTTCACCAAGAGCGTGTGGGGGGCGTCGCAAACATCGGGGGATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3461 ATGAAGTAGCACTGGGTTCACCAAGAGCGTGGGGGGCGTC 3520

Qy   1021 GGCAACAAACACTTTGATCTGCCCCACGGACTGCTTCCGGGAAGCATCCCGGACGGCCACTTAC 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3521 GGCAATAAAAACCTTGACCTGCCCCACGGACTGCTTCCGGGAAGCACCCCGGACGGCCACTTAC 3580

```
Qy    1081  ACCAAAATGCGGGTTCGGGGGCCTTGGTTG 1107
            |||||| || ||||||||||||||||||||
Db    3581  ACCAAGTGTGGTTCGGGGGCCTTGGTTG 3607
```

Search completed: February 21, 2005, 06:24:51
Job time : 709 secs

This Page Blank (uspto)

```
OM nucleic - nucleic search, using sw model

                                    GenCore version 5.1.6
                                    Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:              February 21, 2005, 05:33:47   ; Search time 228 Seconds
                                                     (without alignments)
                                                     7944.554 Million cell updates/sec

Title:               US-09-664-363-19
Perfect score:       1107
Sequence:            1 TCCTCCCGCTGCTGGGTAGC............GCGGGTTCGGGGGCCTTGGTTG 1107
Scoring table:       IDENTITY_NUC
Searched:            Gapop 10.0 , Gapext 1.0

Searched:            1202784 seqs, 81818359 residues

Total number of hits satisfying chosen parameters:        2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
         1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
         2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
         3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
         4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
         5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
         6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1107	100.0	1107	3	US-08-191-160-19	Sequence 19, Appl
2	1107	100.0	2116	3	US-08-191-160-21	Sequence 21, Appl
3	927.8	83.8	11076	4	US-09-539-601-1	Sequence 1, Appli
4	927.8	83.8	11076	4	US-09-539-601-19	Sequence 19, Appli
5	927.8	83.8	11076	4	US-09-539-601-25	Sequence 25, Appli
6	927.8	83.8	11076	4	US-09-539-601-31	Sequence 31, Appli
7	909.8	82.2	9595	3	US-09-014-416-4	Sequence 4, Appl
8	909.8	82.2	9599	3	US-09-014-416-6	Sequence 6, Appl
9	893.2	80.7	9472	4	US-09-204E-96	Sequence 96, Appl
10	891	80.5	9413	4	US-09-827-688-6	Sequence 6, Appli
11	886.2	80.1	6039	2	US-08-324-977-11	Sequence 11, Appl
12	886.2	80.1	6039	2	US-08-324-977-11	Sequence 11, Appl
13	886.2	80.1	6039	3	US-08-384-616-11	Sequence 11, Appl
14	886.2	80.1	6039	3	US-08-904-686A-11	Sequence 11, Appl
15	886.2	80.1	9030	1	US-08-324-977-13	Sequence 13, Appl
16	886.2	80.1	9030	2	US-08-324-977-13	Sequence 13, Appl
17	886.2	80.1	9030	2	US-08-384-616-13	Sequence 13, Appl
18	886.2	80.1	9030	2	US-08-904-686A-13	Sequence 13, Appl
19	886.2	80.1	9030	1	US-09-315-850-13	Sequence 13, Appl
20	886.2	80.1	9416	2	US-09-384-616-1	Sequence 1, Appl
21	886.2	80.1	9416	2	US-08-904-686A-1	Sequence 1, Appli
22	886.2	80.1	9416	3	US-09-315-850-1	Sequence 1, Appli
23	886.2	80.1	9416	3	US-08-823-895A-27	Sequence 27, Appl
24	883	79.8	2082	3	US-08-612-973-47	Sequence 47, Appl
25	883	79.8	2082	3	US-08-927-597-47	Sequence 47, Appl
26	883	79.8	2433	3	US-08-612-973-49	Sequence 49, Appl
27	883	79.8	2433	3	US-08-927-597-49	Sequence 49, Appl
28		62.3	1539	2	US-08-470-426B-17	Sequence 17, Appl
29	689.4	62.3	1863	3	US-08-470-426B-14	Sequence 14, Appl
30	676.6	61.1	1620	4	US-09-194-949A-10	Sequence 10, Appl
31	664.8	60.1	1476	3	US-08-612-973-35	Sequence 35, Appl
32	664.8	60.1	1476	3	US-08-927-597-35	Sequence 35, Appl
33	642.6	58.0	9646	3	US-08-811-566-1	Sequence 1, Appl
34	642.6	58.0	9646	3	US-09-034-756-1	Sequence 1, Appl
35	642.6	58.0	12980	3	US-09-811-566-5	Sequence 5, Appl
36	642.6	58.0	12980	3	US-09-034-756-5	Sequence 5, Appl
37	639.4	57.8	1207	1	US-08-325-630-1	Sequence 1, Appli
38	639.4	57.8	1207	1	US-09-014-416-2	Sequence 2, Appli
39	639.4	57.5	9599	3	US-09-827-688-7	Sequence 7, Appli
40	636.2	57.5	9365	4	US-09-460-806-1	Sequence 1, Appli
41	634.6	57.3	9401	2	US-08-432-693-1	Sequence 1, Appli
42	634.6	57.3	9416	2	US-08-811-566-19	Sequence 19, Appli
43	634.6	57.3	9416	2	US-09-034-756-19	Sequence 19, Appli
44	633	57.2	1207	1	US-08-460-806-3	Sequence 3, Appli
45	633	57.2	1207	1	US-08-325-630-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-08-191-160-19
; Sequence 19, Application US/08191160
; Patent No. 6210675
; GENERAL INFORMATION:
;  APPLICANT: Highfield, Peter Edmund
;  APPLICANT: Rodgers, Brian Colin
;  APPLICANT: Tedder, Richard Seton
;  APPLICANT: Barbara, John Anthony James
;  TITLE OF INVENTION: Viral Agent
;  NUMBER OF SEQUENCES: 25
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Rothwell, Figg, Ernst & Kurz
;  STREET: 1700 K Street
;  CITY: Washington
;  STATE: D.C.
;  COUNTRY: U.S.A.
;  ZIP: 20006
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage
;  COMPUTER: IBM AT compatible
;  OPERATING SYSTEM: MS-DOS V3.2
;  SOFTWARE: Wordperfect 5.0 (DOS text)
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/191,160
;  FILING DATE:
; PRIOR APPLICATION DATA:
;  CLASSIFICATION:
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 07/628,516
;  FILING DATE: 17 DEC 1990
;  APPLICATION NUMBER:
;  FILING DATE: 18 DEC 1989
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: UK 89 28 562.1
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: UK 90 04 414.0
;  FILING DATE: 27 FEB 1990
; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: UK 90 04 814.1
;  FILING DATE: 03 MAR 1990
; ATTORNEY/AGENT INFORMATION:
;  NAME: E. Anthony Figg
;  REGISTRATION NUMBER: 27,195
;  REFERENCE/DOCKET NUMBER: 1645-103A
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (202) 833-5740
;  TELEFAX: (202) 833-5744
; INFORMATION FOR SEQ ID NO: 19:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 1107 base pairs
;  TYPE: nucleotide with corresponding protein
;  STRANDEDNESS: single
```

```
;  TOPOLOGY:  linear
;  MOLECULE TYPE:  cDNA to genomic RNA
;  ORIGINAL SOURCE:
;  ORGANISM:  human; serum infectious for PT-NANBH
;  IMMEDIATE SOURCE:
;  LIBRARY:  clone 136/155
;  FEATURE:
;  LOCATION:  from 1 to 1107 bp portion of the PT-NANBH
;  LOCATION:  polyprotein
;  OTHER INFORMATION:  probably encodes viral structural
;  OTHER INFORMATION:  proteins
US-08-191-160-19

Query Match          100.0%;  Score 1107;  DB 3;  Length 1107;
Best Local Similarity  100.0%;  Pred. No. 1.6e-305;
Matches 1107;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  TCCTTCCCGCTGCTCGGGTAGGCTCACTCCACTCCCACGCGCCCCACGCCCAGCATCCC    60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  TCCTTCCCGCTGCTCGGGTAGGCTCACTCCACTCCCACGCGCCCCACGCCCAGCATCCC    60

Qy     61  ACTCCGACAAATACCGACGCCACGTGCGATTTGCTCGTTCGGTTGGGGGCCTTCTGCTCG   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ACTCCGACAAATACCGACGCCACGTGCGATTTGCTCGTTCGGTTGGGGGCCTTCTGCTCG   120

Qy    121  ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTTCTCAGCTGTTCACCTTCTCG   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  ATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTCGTTCTCAGCTGTTCACCTTCTCG   180

Qy    181  CCTCGGCCGACATCAGAGACGGTACAGGAGACTGCAATGTTCAAATCTATCCCGGCACGTATCCA   240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  CCTCGGCCGACATCAGAGACGGTACAGGAGACTGCAATGTTCAAATCTATCCCGGCACGTATCCA   240

Qy    241  GGTCACCGGCATGGCTTGGGGATATGGATGATGAATGGTCACCTACAGCAGCCCTAGTGGTA   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GGTCACCGGCATGGCTTGGGGATATGGATGATGAATGGTCACCTACAGCAGCCCTAGTGGTA   300

Qy    301  TCGCAGCTACTCCCGGATCCCACAAGCTGTCGTCGGTGGCGGGGGCCCACTGGGGGA   360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  TCGCAGCTACTCCCGGATCCCACAAGCTGTCGTCGGTGGCGGGGGCCCACTGGGGGA   360

Qy    361  GTCCTGGCGGGCGGGCCTTGCCTACTAATTCCATGGTGGGGGAACTGGGCGTTAAGGTCTTGGTTGGTTGTG   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GTCCTGGCGGGCGGGCCTTGCCTACTAATTCCATGGTGGGGGAACTGGGCGTTAAGGTCTTGGTTGGTTGTG   420

Qy    421  ATGCTACTCTTTGCCGGCGTTGACGGGGGAACCTTACACGACAGCAGGGGGGCACAGGGCCGC   480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  ATGCTACTCTTTGCCGGCGTTGACGGGGGAACCTTACACGACAGCAGGGGGGCACAGGGCCGC   480

Qy    481  GCCGCCCACGGGCTTACATCCCTCTTCACACCTGGGGCGCGCTCAGAAAATCCAGCTTTGTTA   540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GCCGCCCACGGGCTTACATCCCTCTTCACACCTGGGGCGCGCTCAGAAAATCCAGCTTTGTTA   540

Qy    541  AACACCAACGGCAGCAGTGGCACACATCAACAGAACTGCCTTGAAACTGAGCTTCCCTCCAA   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  AACACCAACGGCAGCAGTGGCACACATCAACAGAACTGCCTTGAAACTGAGCTTCCCTCCAA   600

Qy    601  ACTGGGGTTCCTTGCCGGCGTGTTTCTACACCGACCACAGGTTCAAATGCCTTCCGGAATGCTCCGGAG   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  ACTGGGGTTCCTTGCCGGCGTGTTTCTACACCGACCACAGGTTCAAATGCCTTCCGGAATGCTCCGGAG   660

Qy    661  CGCAGGGCCAGCTGCGCGCCCCATTGACCCAGTTCGATCAGGGGTGGGGTCCCATCACTTAT   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CGCAGGGCCAGCTGCGCGCCCCATTGACCCAGTTCGATCAGGGGTGGGGTCCCATCACTTAT   720

Qy    721  AATGAGTCCCACGGTTGGACCAGAGGAGGCCCCTTATTGCTGGCACTACCACCTCACCGTGT   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  AATGAGTCCCACGGTTGGACCAGAGGAGGCCCCTTATTGCTGGCACTACCACCTCACCGTGT   780

Qy    781  GGTATCGTGCCCGCGGCGTTGCAGGTGTGTGGGCCCCAGTGTACTGTTCACTCCAAGCCCTGTT   840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  GGTATCGTGCCCGCGGCGTTGCAGGTGTGTGGGCCCCAGTGTACTGTTCACTCCAAGCCCTGTT   840

Qy    841  GTGGTGGGGGACGACCGATCGTCTTCGGCGCGCCTAGCTACTACAGATGGGTTGAGAATGAGACG   900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  GTGGTGGGGGACGACCGATCGTCTTCGGCGCGCCTAGCTACTACAGATGGGTTGAGAATGAGACG   900

Qy    901  GACGTGCTGCTTCTCAACAACACCAGCGGGCGCCCCGGGCAACTGGTTCGGTTCGCGTGTTGTTACATGG   960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GACGTGCTGCTTCTCAACAACACCAGCGGGCGCCCCGGGCAACTGGTTCGGTTCGCGTGTTGTTACATGG   960

Qy    961  ATGAATAGCACCGGTTCACCCAAGACGGTGTGGGGGCCCCCGTCGCAACATCGGGGGGTC   1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  ATGAATAGCACCGGTTCACCCAAGACGGTGTGGGGGCCCCCGTCGCAACATCGGGGGGTC   1020

Qy   1021  GGCAACAACACTTTGATCTGCCCCACGACTGCTTCCGGAAGCATCCGAGGGCCCACTTAC   1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  GGCAACAACACTTTGATCTGCCCCACGACTGCTTCCGGAAGCATCCGAGGGCCCACTTAC   1080

Qy   1081  ACCAAATGCGGTTCGGGGCGCCTTGGTTG   1107
           ||||||||||||||||||||||||||||
Db   1081  ACCAAATGCGGTTCGGGGCGCCTTGGTTG   1107
```

RESULT 2
US-08-191-160-21

```
;  Sequence 21, Application US/08191160
;  Patent No. 6210675
;  GENERAL INFORMATION:
;  APPLICANT:  Highfield, Peter Edmund
;  APPLICANT:  Rodgers, Brian Colin
;  APPLICANT:  Tedder, Richard Seton
;  APPLICANT:  Barbara, John Anthony James
;  TITLE OF INVENTION:  Viral Agent
;  NUMBER OF SEQUENCES:  25
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE:  Rothwell, Figg, Ernst & Kurz
;  STREET:  1700 K Street
;  CITY:  Washington
;  STATE:  D.C.
;  COUNTRY:  U.S.A.
;  ZIP:  20006
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE:  Floppy diskette, 5.25 inch, 360 Kb storage
;  COMPUTER:  IBM AT compatible
;  OPERATING SYSTEM:  MS-DOS V3.2
;  SOFTWARE:  Wordperfect 5.0 (DOS text)
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER:  US/08/191,160
;  FILING DATE:
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:  07/628,516
;  FILING DATE:
;  APPLICATION NUMBER:  UK 89 28 562.1
;  FILING DATE:  18 DEC 1989
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:  UK 90 04 414.0
;  FILING DATE:  27 FEB 1990
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:  UK 90 04 814.1
;  FILING DATE:  03 MAR 1990
;  ATTORNEY/AGENT INFORMATION:
;  NAME:  E. Anthony Figg
;  REGISTRATION NUMBER:  27,195
;  REFERENCE/DOCKET NUMBER:  1645-103A
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE:  (202) 833-5740
;  TELEFAX:  (202) 833-5744
;  INFORMATION FOR SEQ ID NO: 21:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH:  2116 base pairs
;  TYPE:  nucleotide with corresponding protein
;  STRANDEDNESS:  single
;  TOPOLOGY:  linear
MOLECULE TYPE:  cDNA to genomic RNA
```

; ORIGINAL SOURCE:
;   ORGANISM: human; serum infectious for PT-NANBH
;   IMMEDIATE SOURCE:
;   LIBRARY: cDNA clones from 5' end of the genome
;   FEATURE:
;   LOCATION: from 308 to 2116 bp start of the PT-NANBH
;   LOCATION: polyprotein
;   OTHER INFORMATION: viral structural and non-structural
;   OTHER INFORMATION: proteins
US-08-191-160-21

Query Match   100.0%; Score 1107; DB 3; Length 2116;
Best Local Similarity   100.0%; Pred. No. 2e-305;
Matches 1107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TCCTCCCCGCTGCTGGGTACGGTCCACTCCACGCTTCCACGTCCCAGCATCCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 TCCTCCCCGCTGCTGGGTAGGGCTCACTCCACGCTTCCACGTCCCAGCATCCCC 1069

Qy   61 ACTGCGACAATACGACGCCACGTCGATTTGGCTTGCTTCTGCTTCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 ACTGCGACAATACGACGCCACGTCGATTTGGCTTGCTTCTGCTTCTCG 1129

Qy   121 ATGTACGTGGGGGATCTCTGCGGGATCTGTTTTTCCTCGTTGGTTCACCTTCTCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 ATGTACGTGGGGGATCTCTGCGGGATCTGTTTTTCCTCGTTGGTTCACCTTCTCG 1189

Qy   181 CCTCGCCGACATCAGACGGTACAGGACTCCAAATTGTTCAAATCATACCCGGCCACGTATCA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 CCTCGCCGACATCAGACGGTACAGGACTCCAAATTGTTCAAATCATACCCGGCCACGTATCA 1249

Qy   241 GGTCACCCGGCCATGGCTTGGGAATATGATGATGAACTGGTCACCTACAGCAGCAGCCCTAGTGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 GGTCACCCGGCCATGGCTTGGGAATATGATGATGAACTGGTCACCTACAGCAGCAGCCCTAGTGGTA 1309

Qy   301 TCGGCAGCTACTCCGGATCCCACAAGCTGTCGTGTGGACATGTGGCGGGGGCCCACTGGGGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 TCGGCAGCTACTCCGGATCCCACAAGCTGTCGTGTGGACATGTGGCGGGGGCCCACTGGGGGA 1369

Qy   361 GTCCTGGGCGGGCCCTTGCCTCCTACTAATTCCATGGGGGGACTTGGGGGTGGGACATGGTG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 GTCCTGGGCGGGCCCTTGCCTCCTACTAATTCCATGGGGGGACTTGGGGGTGGGACATGGTG 1429

Qy   421 ATGCTACTCTTTGCCGGCGTTGACGGGGAACCTTACACGACAGGGGGGGACACACGGGCCGC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1430 ATGCTACTCTTTGCCGGCGTTGACGGGGAACCTTACACGACAGGGGGGGACACACGGGCCGC 1489

Qy   481 GCCCCCACGGGCTTACATCCCTCTTCACACCTGGGCCGCGCTCAGAAAATCCAGCTTGTA 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1490 GCCCCCACGGGCTTACATCCCTCTTCACACCTGGGCCGCGCTCAGAAAATCCAGCTTGTA 1549

Qy   541 AACACCAACGGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAATGAACTCCCTCCAA 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 AACACCAACGGGCAGCTGGCACATCAACAGAACTGCCTTGAACTGCAATGAACTCCCTCCAA 1609

Qy   601 ACTGGGGTTCCTTGCCGGCGTGTTCTACACCGCACAGGTTCAATGCCTCCGGATGCTCAGAG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 ACTGGGGTTCCTTGCCGGCGTGTTCTACACCGCACAGGTTCAATGCCTCCGGATGCTCAGAG 1669

Qy   661 CGCATGGCAGCTGCCGCGCCCCATTGACCAGTTCGATTCAGGGGTCCCCATCACTACTTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 CGCATGGCAGCTGCCGCGCCCCATTGACCAGTTCGATTCAGGGGTCCCCATCACTACTTAT 1729

Qy   721 AATGAGTCCCACGGCCTTGGACCGACCACTACGCACGCACTCAACCGTGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1730 AATGAGTCCCACGGCCTTGGACCGACCACTACGCACGCACTCAACCGTGT 1789

Qy   781 GGTATCGTGCCCGGCCTTGCAGGGGTGTGTGTGGCCCAGTGTACTCGTTTCACTCCAAGCCCTGTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1790 GGTATCGTGCCCGGCCTTGCAGGGGTGTGTGTGGCCCAGTGTACTCGTTTCACTCCAAGCCCTGTT 1849

Qy   841 GTGGTGGGGGACGACGGATCGTTTCGGCGCCCCTAGTACAGATAGGGGGTGAGAATGAGACG 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1850 GTGGTGGGGGACGACGGATCGTTTCGGCGCCCCTAGTACAGATGGGGGTGAGAATGAGACG 1909

Qy   901 GACGTGCTGCTTCTTCAACAACACGCGGCGCCCGCCGCCACGGGCCAACTGGTTCGGCTACATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1910 GACGTGCTGCTTCTTCAACAACACGCGGCGCCCGCCGCCACGGGCCAACTGGTTCGGCTACATGG 1969

Qy   961 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGGCGCCCGTTGCAAACATCGGGGGGGTC 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1970 ATGAATAGCACCGGGGTTCACCAAGACGTGTGGGGGCGCCCGTTGCAAACATCGGGGGGGTC 2029

Qy   1021 GGCAACAACACTTTTGATCTTCGCCCCACGGACTTGGCTTCCGGGAGGCATCCGAGGCCACTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2030 GGCAACAACACTTTTGATCTTCGCCCCACGGACTTGGCTTCCGGGAGGCATCCGAGGCCACTTAC 2089

Qy   1081 ACCAAAATGCGGTTCGGGGGCCTTGGGTTG 1107
              ||||||||||||||||||||||||||||||
Db   2090 ACCAAAATGCGGTTCGGGGGCCTTGGGTTG 2116

RESULT 3
US-09-539-601-1
; Sequence 1, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;   APPLICANT: Bartenschlager, Ralf FW
;   TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;   FILE REFERENCE: all sequences
;   CURRENT APPLICATION NUMBER: US/09/539,601C
;   CURRENT FILING DATE: 2001-08-30
;   EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;   EARLIER FILING DATE: 1999-04-03
;   NUMBER OF SEQ ID NOS: 51
;   SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11076
;   TYPE: DNA
;   ORGANISM: Hepatitis C virus
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (1)..(341)
;   OTHER INFORMATION: construct I389/Core-3'/wt
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (342)..(1193)
;   OTHER INFORMATION: hepatitis C virus core - neomycin
;   OTHER INFORMATION: phosphotransferase fusion protein
;   FEATURE:
;   NAME/KEY: RBS
;   LOCATION: (1202)..(1812)
;   OTHER INFORMATION: internal ribosome entry site from
;   OTHER INFORMATION: encephalomyocarditis virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1813)..(10845)
;   OTHER INFORMATION: hepatitis C virus polyprotein from core to
;   OTHER INFORMATION: nonstructural protein NS5B; parental sequence
;   OTHER INFORMATION: without cell culture-adaptive mutations
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: (10846)..(11076)
US-09-539-601-1

Query Match   83.8%; Score 927.8; DB 4; Length 11076;
Best Local Similarity   89.9%; Pred. No. 4.6e-254;
Matches 995; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy   1 TCCTCCCCGCTGCTGGGTAGGGCTCACTCCACGCTTCCACGTCCCAGCATCCCC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2515 TCCTCCCCGCTGCTGGGTAGGGCTCACTCCACGCTCCACGCTCCCAGCATCCCC 2574

Qy   61 ACTGCGGACAATACGACGCCACGTCGATTTGCTCGTTGGGTGGGCGGCGCTGGCCTTGTCTCGTCCGCT 120
              ||| ||||||||||||||||||||||| ||||| ||||||||| ||| ||| ||||||
Db   2575 ACTACGACGAATACGACGCCANTGCATTTGCATTGCTCGTTGGGCGGGCGGCGCTGGCCTTGTCTCGTCCGCT 2634

RESULT 4
US-09-539-601-19
; Sequence 19, Application US/09539601C

; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus cell culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 11076
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/Core-3'/9-13F
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION: internal ribosome entry site from
; OTHER INFORMATION: encephalomyocarditis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(10845)
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: hepatitis C virus nonstructural protein NS5B; carries cell
; OTHER INFORMATION: culture-adaptive mutations from clone 9-13F
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)
; US-09-539-601-19

Query Match   83.8%; Score 927.8; DB 4; Length 11076;
Best Local Similarity   89.9%; Pred. No. 4.6e-254;
Matches 995; Conservative   0; Mismatches 112; Indels   0; Gaps   0;

Qy   1 TCCTCCCGCTGCTGGGGTAGGCTGCTCACTCCCACACGCTTCGCGCCGCCAAGGACATCCCC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2515 TCCTCCCGCTGCTGGGTAGCCCTCACTCCCACGCCTCGCGCCGCCCAGGACGCATCGTCC 2574

Qy   61 ACTGGGACAATACGACGCCACGTCGAATTGCTCGTTCGTTGCCCTTCGTCCGGCTCCGCGCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2575 ACTACGACGATACGACCGCCATGTCGAATTGGCTCGTTCGGGGGCTGCTCTCTGCTCCGGCT 2634

Qy   121 ATGTACGTGGGGGGATCTCTGCGGATCTGTTTTCCTTCGTCTCTCAGCTGTTCACTTTCTCG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2635 ATGTACGTGGGGAGAATCTCTGCGGATCGTGTTTTCCTTGCGCCCAGCTGTTCACTTTCTCG 2694

Qy   181 CCTCGGCGGACATCAGACGGTACAGGAGGACTGGCAATTGTTCAATCTCATCTCCGGCCTATCA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2695 CCTCGGCCGGACAGAGACAGTACACGGACTGCAATTGCTCAATATATCCCGGACAGTGACA 2754

Qy   241 GGTCACCGCCATGGCGCTTGGGGATATGATGATGAACTGGTCACCTACAGCAGCCCCTAGTGGTA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2755 GGTCACCGTATGGCTTGGGGATATGATGATGAACTGGTCACCTACAGCAGCCCCTAGTGGTA 2814

Qy   301 TCGGCAGCTACTCCGGATCCCACAACAAGCTGTCTGTTGGACATGGTGGGGGGGGGCCCACTGGGGA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2815 TCGGCAGTTACTCCGGATCCCACAAGCTGTCGTTGTGGGATATGGTGGGGGGGGCCCCACTGGGGGA 2874

Qy   361 GTCCTGGCGGCGGCCTTGCCTACTACTATTCCATGGTGGGGGGACTGGGGCTAAGGTTCTTGGTTGTG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2875 GTCCTAGCGGCGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGGCTAAGGTTCTGATTGTG 2934

Qy   421 ATGCTACTCTTTGCCGGGGCGTTGACGGGGGAACCTTACACGACAGCAGGGGACACGGCCGCGC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2935 ATGCTACTCTTTGCCGGGGTTGCACGGTTCTTACACGGACAGGGGGAACGATGGCCAAA 2994

Qy   481 GCCGCCCACGGGCGTTACATCCCTCTTCACACCTGGGCCGGCGGCTCAGAGAAAATCCAGCTTGTA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2995 AACACCCTCGGGATTACGTCCCTGTTTTTCACCCGGGTCATCCCAGAAAATCCAGCTTGTA 3054

Qy   541 AACACCAACGGCAGCTGGCGACATCAACAGAACTGCCTTGAACTGCAATGACTCCCTCCAA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3055 AACACCAACGGACAGTGGCGCCATCAACATGGCTGGCGAATGACTCGCCTCTCAC 3114

Qy   601 ACTGGGTTCCTTGCCGGGTGTTCTTACACGGCACAGGTTCAATGGCTGCTCAGAG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3115 ACTGGGTTCCTTGCTGCTGTTCTACGTGCACAAGTTCAACTCATCTGGATGCCCCAGAG 3174

Qy   661 CGCATGGCCAGCTGCCGCGCCCCATTGACCAGTTCGATCAGCGGGTGGGGTCCCATCACTTAT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3175 CGCATGGCCAGCTTGCGAGCGCCCCATCGACGCGTTCGCTGCTCAGGGGTGTGGGGTCCCATCACTTAC 3234

Qy   721 AATGAGTCACCAGCCTTGGACCAGAGGCCCTATTGCTGGCACTAGCGACACGCACCCGGCCGGTG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3235 AATGAGTCACACAGCTCGGACCAGAGGCCTTATTGTTGGCACTAGCGACACGCCCGCGGGTGGC 3294

Qy   781 GGTATCGTGCCCGGGGTTGCAGGGTGTGTGGGCCCAGTGTACTGTTTTCACTCCAAGCCCTGTT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3295 GGTATCGTACCGGGGCAGGTGTGTGGGGTTCTACTGCCTTCACACCCAAAGCCCTGTC 3354

Qy   841 GTGGTGGGGACGACGACCGGATCGTTTCGGGCGCCCTACGTACAGACAGTGGGTGAGAATGAGAGCG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3355 GTGGTGTTGGGGACGACGACCGGGTTCGGGCGGTCCTACAGTTGGGGGAGAAATGAGAGACG 3414

Qy   901 GACGTGCTGCTTCTCAACAACACACGGGGCGCGCCCGCCAACTGGTTCGGGCTGTAACATGG 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3415 GACGGCTGCTTCTTAACAACACGCGGGGGCGCGGCCAACTGGTTTGGCTGGTTACATGG 3474

Qy   961 ATGAATAGCACACGGGGGTTCACCAGCAGTGTGGGGGGGCCCCGTTCGAAACATCGGGGGGGATC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3475 ATGAATAGCACTGGGGTTCACCAGCACGGTGGGGGTGCTGTGAAACACTCGGGGGGTC 3534

Qy   1021 GGCAACAACACTTTTGAATCTGCCCCCGGAAGCAATCCCGAGGCGCCACTTAC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3535 GGCAATAAAACCTTGACCTGCCCCACGGACTGCTTCCCGGAAGCACCCCGAGGCCCACTTAC 3594

Qy   1081 ACCAAATGCGGGTTCGGGGGCCTTGGTTG 1107
             ||||||||||||||||||||||||||||
Db   3595 ACCAAGTGTGTGGTTCGGGGGCCTTGGTTG 3621

RESULT 5
US-09-539-601-25
Sequence 25, Application US/09539601C
Patent No. 6630343

GENERAL INFORMATION:

; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 25
; LENGTH: 11076
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/Core-3'/5.1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)

; OTHER INFORMATION: hepatitis C virus core - neomycin
; OTHER INFORMATION: phosphotransferase fusion protein
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION: internal ribosome entry site from
; OTHER INFORMATION: encephalomyocarditis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(10845)
; OTHER INFORMATION: hepatitis C virus polyprotein from core to
; OTHER INFORMATION: nonstructural protein NS5B; carries cell
; OTHER INFORMATION: culture-adaptive mutations of clone 5.1
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (10846)..(11076)

US-09-539-601-25

Query Match   83.8%;   Score 927.8;   DB 4;   Length 11076;
Best Local Similarity   89.9%;   Pred. No. 4.6e-254;
Matches 995;   Conservative   0;   Mismatches 112;   Indels   0;   Gaps   0;

[Sequence alignment of Query (Qy) and Database (Db) rows follows, comparing nucleotide positions]

RESULT 6

US-09-539-601-31

Sequence 31, Application US/09539601C

Patent No. 6630343

GENERAL INFORMATION:

APPLICANT: Bartenschlager, Ralf FW

TITLE OF INVENTION: Hepatitis C Virus Cell Culture System

FILE REFERENCE: all sequences

CURRENT APPLICATION NUMBER: US/09/539,601C

CURRENT FILING DATE: 2001-08-30

EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY

EARLIER FILING DATE: 1999-04-03

NUMBER OF SEQ ID NOS: 51

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 31

LENGTH: 11076

TYPE: DNA

ORGANISM: Hepatitis C virus

FEATURE:

NAME/KEY: 5'UTR

LOCATION: (1)..(341)

OTHER INFORMATION: construct I389/Core-3'/19

FEATURE:

NAME/KEY: CDS

LOCATION: (342)..(1193)

OTHER INFORMATION: hepatitis C virus core - neomycin

OTHER INFORMATION: phosphotransferase fusion protein

FEATURE:

NAME/KEY: RBS

LOCATION: (1202)..(1812)

OTHER INFORMATION: internal ribosome entry site from

OTHER INFORMATION: encephalomyocarditis virus

FEATURE:

NAME/KEY: CDS

LOCATION: (1813)..(10845)

OTHER INFORMATION: hepatitis C virus polyprotein from core to

OTHER INFORMATION: nonstructural protein NS5B; carries cell culture

OTHER INFORMATION: adaptive mutations from clone no. 19

FEATURE:

NAME/KEY: 3'UTR

LOCATION: (10846)..(11076)

US-09-539-601-31

Query Match   83.8%;   Score 927.8;   DB 4;   Length 11076;

Best Local Similarity   89.9%;   Pred. No. 4.6e-254;

Matches 995; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

RESULT 7
US-09-014-416-4
; Sequence 4, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 9595
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-4

Query Match   82.2%; Score 909.8; DB 3; Length 9595;
Best Local Similarity   89.0%; Pred. No. 5.9e-249;
Matches 983; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy   3 CTCCCGCTCGTGGGTAGCCGCTCACTCCCACGCTCGCGGCCAAGGACGCAGCATCCCAC 62
           ||||| ||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Db   1046 CTCCCGTTCCTGGGTAGCCGCTCACTCCCACGGTTCCCCAGGAATGCCAGGGTCCCCAC 1105

Qy   63 TGCCACAAATACGACGCCCACGTCGATTTGTCGTTGCGTGGGGGGGTGGGGGTCGCGTAT 122
           |||| ||||||||||||||||| |||| ||||||| |||||||||||| ||||| |||||
Db   1106 TACGACAATACGACGCCCAGGTCGACTTGCTCGTTGTTGGGACGGCTGCTTTTTCTCCGCTAT 1165

Qy   123 GTACGTGGGGGGATCTCGGGGATCTGTTTTTCCTTCGTCTTCAGCTGTTCACCTTCTCGCC 182
           ||||| |||||||||||| |||||| ||||||| |||||||||||||| |||||||||||
Db   1166 GTACGTGGGGGGATCTCTCGGGATCTATTTTTCCTCGTCTCCCAGCTGTTCACCTTCTCGCC 1225

Qy   183 TCGGCCGACATCAGACGGTACAGGACTGCAATTGTTCAATCTATCCCAGG 242
           ||||| ||||||||||||| |||||||||||||||||| ||||||||||||
Db   1226 TCGGCCCGCATGAGACAGTGCAGACTGCGAACTGCTCAATCTATCCCAGG 1285

Qy   243 TCACCGCATGGCTTGGGATATGATGATGAACTGTCACCTCACGCAGCCTAGTGGTATC 302
           ||||| |||||||||||| ||||||||| || ||||||| |||||||||||||||||||
Db   1286 TCACCGCATGGCTTGGGATATGAACTGGTCACCTACACAGCCCTTAGTGGTGTC 1345

Qy   303 GCAGCTACTCCGGAGATCCCACAAGCTGTGTCCGTTGGGACATGGTTGGGGAGT 362
           |||||||||||||| ||||||||||||||| ||||||| |||||||||||||||
Db   1346 GCAGTTGCTCCCGGATCCCACAAGCTGTGTCCTGTTGGGGACATGGTTGGGGAGT 1405

Qy   363 CCTGGCGGCGCCCTTCCCTACTATTCCATGGTGGGGAACTGGGTCTTGGTTGTTGTGAT 422
           |||||| ||||||||||| || ||||||| ||||||||| ||||||| || |||||||
Db   1406 CCTGGCGGGGCCCTTGCCCTACTATTCCATGGTGGGGAACTGGGTTCTTGATTGTGTGGC 1465

Qy   423 GCTACTCTTTGCGGGCGTTGACGGGGAACCTTACGGACGACGAGGACACACGGCCGCGCC 482
           |||| ||||| ||||||| ||||| ||||||||||||||| ||||||||||||||||||||
Db   1466 GCTACTCTTTGCCCGCGTTGACGGGGAGACCCACACGGACGGCCACAC 1525

Qy   483 CGCCCACGGGCTTACATCCCTCTTCACACCTGGGGCCGCGGGCTCAGAGAAAATCCAGCTTGTAAA 542
           ||||||| |||||||||||||||||||| | ||| ||||| |||||||||||||||||||||||||
Db   1526 CACCTCCCGGGTTCACGTCCCTTTCTCATCTCGGACGTCAGAAAATCCAGCTTGTGAA 1585

Qy   543 CACCAAGCGCAGCTGGCACATCAACAGACTGCTTGACTGCAATGACTCCCTCCAAAC 602
           |||||||||| ||||| ||||||||| ||||||||||| |||| ||||||||||||||
Db   1586 TACCAAACGGCCAGTGGGACACATCAACAGAGGGACTGCTGCTTTGGGACAACTCCCTCCAAAC 1645

Qy   603 TGGGTTCCTTGCGCGGCGCGTGTTCACGCGACAGTTCAATGCGTCCGGATGCTCAGACGCG 662
           |||||||| ||| ||||||||||||||| || ||||| |||||||||| || |||| ||||
Db   1646 TGGGTTCTTTGCCGGCGCCGTGTTTACGGCACAAGTTCAACTCGTCCGGGGTGGGGTCCCGGAGCG 1705

Qy   (continued)
Db   3595 ACCAAGTGTGGTTCGGGGGCCTTGGTTG 3621
           |||||||||| ||||||||||||||||||
Db   (matching line)

Db   1706 CATGGGCCAGCTGCCGCCGCCCATTGACCAGTTCGATCAGGGGGTGGGGGTCCCATCACACTTATAA 1765

Qy   663 CATGGGCCAGCTGCCGCCGCCCATTGACCAGTTCGATCAGGGGGTGGGGGTCCCATCACACTTATA 722

Db   1766 TAAGCCTAACAGCTCGATCAGAGAGGCCCTTATTGCTGGGACTTACGGCCGCCTCGACCCGTGTGG 1825

Qy   723 TGAGCGTCCCACGGCTTGGGACCGCCCCATTGACTGGTTCGCCCGCCACAGGGGTGGGGGCCCCATCACCTCACCTATAC 782

Db   1826 TGTCGTACCCCGGCTCCCAGGTGTGTGGGTCCGGTGTGTGGGTCCAGTGTATATTGTTTCACCCGGAGGCCCCTGTTGT 1885

Qy   783 TATCGTGCCCCGCGTTGCAGGTGTGGGGCCAGAGTAGTACGTTCACTCCAAGCCCTGTTGT 842

Db   1886 GGTGGGGGACCAACGCAGCGTTCCGGTGTCCCTACGGTATAGCTGGGGGGGAGGAATGAGACGA 1945

Qy   843 GGTGGGGGACGACGAGATCGTTTTTCGGGGCGCCCCCTTACGGTACAGATGGGGGTGAGAATGAGACGGA 902

Db   1946 CGTGATGCTCCTCAACAACACGCGTCCGGCCCACAAGGCAACTGGGTTCGGGCTGTACATGGAT 2005

Qy   903 CGTGCTGCTTCTCAACAACACGCGGCCGCCCGGGGACAACTGGGTTCTGGCTGTACATGGAT 962

Db   2006 GAATAGTACTGGGGTTCACTAAGGACGTGCGGAGGGTCCCCCGTTGTAACATCGGGGGGTCGG 2065

Qy   963 GAATAGCACCGGGGTTCACCAAGACGGTGTGGGGGCCCCCCGTTGCAACATCGGGGGGTCGG 1022

Db   2066 TAACCCGCACCTTGATCTGCCCCACGGACTGCTTCCGGAAGCACCCCGGAGGCTACTTTACAC 2125

Qy   1023 CAACAACACTTTGATCTGCCCCACGGACTCGCTTCCGGAAGGCATCCCAGGCCCGGACATTCCCAC 1082

Db   2126 AAAAATGTGGGTCGGGGGGCCCTTGGTTG 2150

Qy   1083 CAAATCGGTTCGGGGGCCCTTGGTTG 1107

RESULT 8
US-09-014-416-6
; Sequence 6, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 9599
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-6

Query Match   82.2%; Score 909.8; DB 3; Length 9599;
Best Local Similarity   89.0%; Pred. No. 5.9e-249;
Matches 983; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy   3 CTCCCGCTCGTGGGTAGCCGCTCACTCCCACGCTCGCGGCCAAGGACGCAGCATCCCAC 62
           ||||| ||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Db   1046 CTCCCGTTCCTGGGTAGCCGCTCACTCCCACGGTTCCCCAGGAATGCCAGGGTCCCCAC 1105

Qy   63 TGCCACAAATACGACGCCCACGTCGATTTGTCGTTGCGTGGGGGGGTGGGGGTCGCGTAT 122
           |||| ||||||||||||||||| |||| ||||||| |||||||||||| ||||| |||||
Db   1106 TACGACAATACGACGCCCAGGTCGACTTGCTCGTTGTTGGGACGGCTGCTTTTTCTCCGCTAT 1165

Qy   123 GTACGTGGGGGGATCTCGGGGATCTGTTTTTCCTTCGTCTTCAGCTGTTCACCTTCTCGCC 182

RESULT 9
US-08-150-204E-96
; Sequence 96, Application US/08150204E
; Patent No. 658126
; GENERAL INFORMATION:

APPLICANT: CHO, Joong Myung
          LEE, Yong Beom
          PARK, Young Woo
          LIM, Kook Jin
          CHOI, Deog Young
          SO, Hong Seob
          KIM, Chun Hyung
          KIM, Sung Taek
          YANG, Jae Young
          YANG, Jae Young
; NUMBER OF SEQUENCES: 128
; TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YANG, Jae Young
; STREET: 386-1, Doryong-dong, Yuseong-gu
; CITY: Daejeon
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 305-340
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage
; COMPUTER: IBM PC/pentium
; OPERATING SYSTEM: Windows
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/150,204E
; FILING DATE: 20-Apr-1994
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 91-9510
; FILING DATE: 10-JUN-1991
; APPLICATION NUMBER: KR 91-13601
; FILING DATE: 6-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Shahan Islam, Esq.
; REGISTRATION NUMBER: 32,507
; REFERENCE/DOCKET NUMBER: 2695/FLK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 940-8564
; TELEFAX: (212) 940-8776
; INFORMATION FOR SEQ ID NO: 96
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9472 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; OTHER INFORMATION: KHCV-LBC1, Fig. 2
; SEQUENCE DESCRIPTION: SEQ ID NO: 96

US-08-150-204E-96

Query Match   80.7%; Score 893.2; DB 4; Length 9472;
Best Local Similarity 88.0%; Pred. No. 3.2e-244;
Matches 973; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

[Sequence alignment between Qy (query) and Db (database) strands follows]

Query Match   80.5%;   Score 891; DB 4; Length 9413;
Best Local Similarity   87.8%;   Pred. No. 1.3e-243;
Matches   972; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy   1 TCCTTCCCGCTGTGGGTAGGCGGTTAGGCTCGCTCACTCCCACGCGTTCGTCCGGCGATCCCC 60
          |||| |||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1032 TTCTCCGGTTGCTGGGGTAGGCGGTTAGGCTCACTCCCACGCGTTCGTCCGGCGCATCCCC 1091

Qy   61 ACTGCGGACAATACGACGCCCACGTCGAGTTGCTGGTTGCTGGTTGGGGGGGGCTGGCCTTT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 ACCACGACAATACGACGCCCACGTCGAGTTTGCTCGTTGGGGGGGGGGGGCTGCTTTCTCGGCT 1151

Qy   121 ATGTAGGTGGGGGGGGACTCTCTGCGGGGATCTGTTTTCTCTCGTTGTCAGCTGTCAGTCTCG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152 AATGTACGTTGGGGGATCTCTGCGGGATCCGTTTTTCTCGTTCTCCCAGCTGTTCACCTTCTCA 1211

Qy   181 CCTTCGGCCGACATCAGAGCGGTACAGGACTGCAATTGGTTCAATCTTATCCCGGCACGTATCA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212 CCTTCGCCGGTTGATGAGACGGGTTACAAGATTGCAAATCTTATCCCGGCCACGTTATCA 1271

Qy   241 GGTCACCGCATGGCTTGGGATATGATGATGAGAACTGGTCACCTTACAGCGACGCCCTAGTGGTA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1272 GGTCACGCCATGGCTTGGGATATGATGATGAACTGGTCACCTTACACTACAACGGCTCACTAGTGGTA 1331

Qy   301 TCGCAGCTACTCCGGGATCCCACACAACGCTGTTCGTCGTGACATAGTGGGCCCACTGGGGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 TCGGCAGCTACTCCCGGGATCCCCAACGGCGTCGCCTGAAGGTTGGCGGCCCCACTGGGGT 1391

Qy   361 GTCCTCGGCGCGGGCCTTTGCCCTACTATTCCATGGTCGGGGGGGGGGGGGGCTAAGGTCTTAAGGTCG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 GTCCTAGCGCGGCGCCTTGCCCTACTATTCCATGGGAACCTGAGGGGGGGGCTAAGGTCTTGATTGTG 1451

Qy   421 ATGGTACTCTTTGCCGGCGTTGACGGGGGAACCTTACACGACAGAGGGGGGGACACACGGCCGC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1452 ATGCTACTCTTTGCTCGGCGGTTGACGGGGCACACCCACGGTGACAGAGGGAAGGTTAGCCTCC 1511

Qy   481 GCCGCCCACCGGGCTTACATCCCTCTTCACACTCGGGGCCGGCCTCAGGAAAATCCAGCTTGTA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512 GCCGCCCACCGGCCTTACATCCCTCTTCACACCTCGGGCCGCGCGTCAGAAAATCCAGTTGTA 1571

Qy   541 AACACCAACGGCACGGCACATCAACACAGAACTGCTTGACTCGTTTGCTCGTTCCCTCCAA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1572 AACACCAACGGCACGTGGCACATCAACACAGAACTGCCTTGAACTGCAATGACTCCCTCCAA 1631

Qy   601 ACTGGGTTTCCTTGCCGCGCGCGCCCCCAGTTCGATCAGGGGGGGGGGGTCCCATCACGTTAT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632 ACTGGGTTCCTTGCCCGCGTTCTACGCCACACCAGGTTCAATGCGTCCGGGATGCTCAGAG 1691

Qy   661 CGCATGGCCAGCTGCGCGCGCCCCATTGACCCAGTTCGATCAGGGGGGGGGTCCCATCACTTAT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692 CGCATGGCCAGCGTGCGCGGCCCCCATTGACCAGTTCGATCAGGGGGGGGTCCCATCACTTAT 1751

Qy   721 AATGAGTCCCACGCGTTGGACCCAGGAGGCCCTTATTGCTGGGCACTACCGCACCGTGT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752 GATATGCCTGAGAGCTCGGACACCAGGAGGCCCATTATTGCTGGGCACTACCGCACCGTGC 1811

Qy   781 ACTGAGCCTCATGATTGGATCAGAGGACGCCCTTATTGCTGGCACTACCGCACCGTCAACCGTGT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1812 GGGATCGTGCCTCGGCCTCGCAGGTGGTGTCCAGTGTATTGCTTCACTCTGTT 1871

Qy   841 GTGGTGGGGGACGACCGGATCGTTTCGGGTGCCCTCCCCGCCAAAATGAGACG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872 GTAGTGGGGACGACCGATCGTTTCGGCGCCTCTACTTATAGCTGGGGGAGAAATGAGACA 1931

Db   1285 GGGTCACCGCATGGCCTGGGATATGATGATGAACTCGTCGCCTACAACAGCCCTAGTGGTA 1344
Qy   301 TCGCAGGCTACTCCGGGATCCCACAAGCTGTCGTGTCGTGACATGGTGGGGGGCCCACTGGGGA 360
Db   1345 TCGGAGCTACTCCGGATCCCACAAGCTGTCGTGTCGTGACATGGTGGGGGGCCCACTGGGGA 1404
Qy   361 GTCCTTGGCCCGGCGCCTTGCCCTACTATTCCATGGTGGGGGGGGGGCTAAGGTCTTGGTTGTG 420
Db   1405 ATCCTGGCCGGGCCTTGCCCTACTATTCCATGGTGGGGGGGGGGGCTAAGGTCTTGGTTGTG 1464
Qy   421 ATGGTACTCTTTGCCGGCGTTGACGGGGGAACCTTACACGACAGAGGGGGGGACACACGGCCGC 480
Db   1465 ATCGTACTCTTTGCCGGCGTTGACGGAACCACCCCACTGACAGGGGGGGCGGCAAGGTCGG 1524
Qy   481 GCCGCCCACCGGGCTTACATCCCTCTTCACACTCGGGGCCGGCCTCAGGAAAATCCAGCTTGTA 540
Db   1525 GCCCCTAGCTCGCTAACGTCCCTCTTAGCCTCGGGGCCGGCCTTCAGCACTCATA 1584
Qy   541 AACACCAACGGCACGGCACATCAACACAGAACTGCTTGACTCGTTTGCTCGTTCCCTCCAA 600
Db   1585 AACACCAACGGCACGGCACTGGCANATATCAACAGGACCGCCCTGACTCGTCGCAATATGACTCCCTCAAC 1644
Qy   601 ACTGGGTTTCCTTGCCGCGCGCGCGCCCCCAGTTCGATCAGGGGGGGGGTCCCATCACGTTAT 660
Db   1645 ACTGGGTTTTGTTGCGCGCTGTTCTTACAAAATACGGTTGCGGTGCCCGGAG 1704
Qy   661 CGCATGGCCAGCTGCGCGCGCCCCATTGACCCAGTTCGATCAGGGGGGGGGTCCCATCACTTAT 720
Db   1705 CGCTTGGCCCACGTTGGACCCCCCATTGATACATTCGGGCGCAGGGGTGGGGTCCCATCACTTAC 1764
Qy   721 AATGAGTCCCACGCGTTGGACCCAGGAGGCCCTTATTGCTGGGCACTACCGCACCGTGT 780
Db   1765 ACTGAGCCTCATGATTGGATCAGAGGACGCCCTATTGCTGGGCACTACCGCACCGTCAACCGTGT 1824
Qy   781 GGTATCGTGCGCCGGGCCTTGCAGGTGTGGGGCGTACTGGTTTCACTTCCAAGCCCTGTT 840
Db   1825 GGTATTGTGCCCAGGTGTGGGCAGGTGTGGGGGGCGTATACTGGCTTCACCCCGAGTCCTGTT 1884
Qy   841 GTGGTGGGGGACGACCGGATCGTTTCGGGTGCCCTCCCCGCCAAAATGAGACG 900
Db   1885 GCGGTGGGGGACTACGGATCGTTTCGGGTGCCCTCCCCGCCAAAATGAGACG 1944
Qy   901 GACGTGCTGCTTCTCAACAACACGCGGCCGCCCACGGGGCAACTGGTTGTTCGGTTACATGG 960
Db   1945 GACTTGCTGCTCTCTTAACAACGCCCGGGCCACTGGTTCGGCTGTACATGG 2004
Qy   961 ATGAATAGGACCGGGGTTCACCAAGACGTGTGTGGGGGGGGCCCCCCCGTGCAACATCGGGGGGGTC 1020
Db   2005 ATGAATGGCACTGGGGTTCACCAAGACATGTGGGGGGGGCCCCCCGTGTAACATCGGGGGGGTC 2064
Qy   1021 GGCACACACACTTTGATCTGTCCCCACGGACTGCTTCCGGAAAGCAATCCCCACAGCTCCTAC 1080
Db   2065 GGCAACAATACCTTGACTGTTCCCCACGGACTGCTTCCGAAAGCACCCCGGGGGGGCCACTTAC 2124
Qy   1081 ACCAAAATCGGGGTTCGGGGGCCTTTGGTT 1106
Db   2125 ACCAAAATGGGGTTCGGGGGGCCTTGGTT 2150

```
Qy   901  GACGTGCTGCTTCTCAACAACACCGGCCCCACGGGGGGGCAAACTTGGTTCCGCGTGTACATGG  960
          ||||||||  |  || |||| || |  ||  ||||||||  |  || ||||| |   | | ||
Db  1932  GACGTGCTCGCTACTTAGCAACACACGCGGGGCCCCTCAAAGGCAACTGGTTGGGTGCACGTGG  1991

Qy   961  ATGGAATAGCACCCGGGTTCACCGACGTGTGGGGGCCCCGCCCCCCGTTGCAACATCGGGGGGGTC  1020
          ||||  |  | |   |||||  ||  |||| ||||| ||||||||||  | ||| ||||||||||
Db  1992  ATGAACAGCACTGGGTTCACCAAGACGTCGCGGGGCCCCTCCGTTGCAACATCCCAAACATCGGGGGGGTC  2051

Qy  1021  GGCAACACACTTTGATCTGCTTCCGGAAGCATCCGGAGGGGCCACTTAC  1080
          ||||  ||| |  |  |  ||||||   |||  | ||| |||||| |||
Db  2052  GGCAACAACACCTTGGTCTGCTGCCCCACGGGATTGCTTCCGGAAGCACCCCGAGGCCACTTAC  2111

Qy  1081  ACCAAAATGCGGTTCGGGGGGCCTTGGTTG  1107
          |||| | |  ||||||| |||  ||||||||
Db  2112  ACAAAGTTGGGCTCGGGGGCCCTGGTTG  2138
```

RESULT 11
US-08-324-977-11
Sequence 11, Application US/08324977
Patent No. 5747339
GENERAL INFORMATION:
  APPLICANT: OKAYAMA, Hiroto
  APPLICANT: FUKE, Isao
  APPLICANT: MORI, Chiaato
  APPLICANT: TAKAMIZAWA, Akahisa
  APPLICANT: YOSHIDA, Iwao
  TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
    cDNA AND ANTIGEN POLYEPTIDE
  NUMBER OF SEQUENCES: 50
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Armstrong, Westerman, Hattori, McLeland &
      Naughton
    STREET: 1725 K St. N.W. Suite 1000
    CITY: Washington
    STATE: D.C.
    COUNTRY: U.S.A.
    ZIP: 20006
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
    SOFTWARE: ASCII
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/324,977
    FILING DATE: 18-OCT-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 2-167466
    FILING DATE: 25-JUN-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 2-230921
    FILING DATE: 31-AUG-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 2-305605
    FILING DATE: 09-NOV-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/099,706
    FILING DATE: 30-JUL-1993
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/769,996
    FILING DATE: 02-OCT-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/635,451
    FILING DATE: 28-DEC-1990
  ATTORNEY/AGENT INFORMATION:
    NAME: Stevens-Smith, Theresa M.
    REGISTRATION NUMBER: 36,281
    REFERENCE/DOCKET NUMBER: 9007030
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (202) 659-2930
    TELEFAX: (202) 887-0357
    TELEX: 440142
  INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
  LENGTH: 6039 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: cDNA from genomic RNA
  FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 1..6039
    OTHER INFORMATION: /note: "sequence = 333 - 6371 of
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..6039
US-08-324-977-11

Query Match 80.1%; Score 886.2; DB 1; Length 6039;
Best Local Similarity 87.5%; Pred. No. 2.7e-242;
Matches 969; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
```

```
Qy     1  TCCTCCCGCTGCTGGGTAGGGCTAGCGCTTCACTCCCACGCTTCGCGGCGGCCAGGAACGTGCAATCCCC  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   703  TCCTCCCGCTGTGCTGGGTAGGGCTAGCGCTCACTCCCACGCTTCGCGGCGGCCAGGAACGTGCAATCCCC  762

Qy    61  ACTGCGACAATACGACGGCCACGTCGAATTTGCTCGTTGGGCGGCGTGCTGCCTTCGTGCTCCGCT  120
          |||||  |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   763  ACCACGACGATACGACGCCCACGTCGAATCGCGTCGTTGGGCGGCGGCGCTTTCGTGCTCCGCT  822

Qy   121  ATGTTACGTTGGGGGATCTCTGCGGATCTGTGTTTTCTCCGTCTCTCAGCTTGTTCACTTCTCG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   823  ATGTTACGTTGGGGGACCTCTGCGGATCTGTTTTCTCGTCTCTCAGCTGTTTCACTTCTCG  882

Qy   181  CCTCGGCCGACATCAGACGCGTACAGGACACTGCAAAGTTGTTCAATCTATTCCCGCGGCCACGTATCA  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   883  CCTCGCCGGGCATTGACATTACAGGACACTGTAACTGCTCAATTTATCCCGGCGGCCCATGTGTCG  942

Qy   241  GGTCACCGCATGGCTTGGGATATGATGAATGGTCACCTACAGCAGCCCTAGTGGGTA  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   943  GGTCACCGTATGGCTTGGGGACATGATGAACTGGTCGCCCACAACAGCGCCCTAGTGGTG  1002

Qy   301  TCGCAGCTACTCCCGGATCCCACAAGCCTGTCGTGGACAGGTAGGTGGCGGGCCCCACTGGGGA  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1003  TCGCAGTTACTCCGGATCCCACAAGCCGTCGTGGACATGGTGGCGGGGCGCCCACTGGGGA  1062

Qy   361  GTCCTGGCGGGGCCCTTGCCTACTATTCCATAGGTGTGGGGGAACTGGGCTTAAGGGTTCTTGGTTGTG  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1063  GTCCTGGCGGGGCCTTGCCTACTATTCCATGGCGGGGAACTGGGCTAAGGGTTCTGATTGTG  1122

Qy   421  ATGCTACTCTTTGCCGGCCGTTGACGGGGAACCTTACACGACGAGGGGGGGGACACGGCCGC  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1123  ATGCTACTTTTTGCTGGCCGTTGACGGGGATACCCACGTGACAGGGGGCGGGGGAAGCCAAA  1182

Qy   481  GCCGCCCACGGGGCTTACATCCCTCCTTCACACCTGGGCCCGGCCTCAGAAAAATCCAGCTTGTA  540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1183  ACACCCAACAGGCCTGTGTCCATGTTCGCAAGTGGGCCGGTCTCAGAAAATCCAGCTTTATA  1242

Qy   541  AACACCAACGGCGACGCTGGGCACATCAACAGAACTGCCTTGAAACTGCAATGACTCCCTCCAA  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1243  AACACCAAATGGGAGTTGGCACATCAACACAGACTGCCCCTGAAACTGCAATGACTCTCTCCAG  1302

Qy   601  ACTGGGTTCCTTGCCGCGCGTTCTCTACACGCACAGGTTCAAATGCGTCCGGATTGCTCAGAG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1303  ACTGGGGTTTTCTTGCCGCGCGTTCTACACGCACAGGTTCAATGCGTCCGGGTGCCCAGAG  1362

Qy   661  CGCATGCGGCCAGCTGCCCGCCCCCATTGACCAGTTCGATCAGGGGGTGGGGGTCCCAGTCACTTAT  720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1363  CGCATCGGGCCCAGTGCGCACCATTGACAAGTTGCGCCAGGGGGTCCCATTACTTAT  1422

Qy   721  AATGAGTCCCACGGCTTGGACCAGAGGCCCTATTGTGTTGGCCACTACGGCACCTCAAACCGTGT  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1423  GCTGAGTCTAGCCAGATCAGACCCAGAGGCCATATATTGCTGGCACTACCCACCAAATGT  1482

Qy   781  GGTATCGTGCCCGCGTTGCAGGTGTGTGGGCCCAGGTACTGGTTTCACTCCAAGCCCTGTT  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ...
```

```
Db  1483  ACCATCGTACCTCGTGCCTTCGCGGAGGTGTTGCCGCGCGGGCCCAGTGTACGGTGTACTTGCTTCACCCGAAGCCCTGTC 1542
               ||||||||||||||||||||||||||||||||||||  |||| ||||||||||||||||
Qy   841  GTGGTGGGGACGACCGATCGTTTCGGCGCGCCCTACGTACAGATGATGGGGGTGGAGAATGAGACG 900

Db  1543  GTCGTGGGGACGACCGATCGTTTCGGTGTCCCTACGTATAGATGGGGGGGGAGAACGAGACT 1602
               |||||||||||||||||||||||
Qy   901  GACCTGCTCGTTCTTCAACAACACACGGGCGCCCCGCCCCGGCGCAACTGGTTCGGCGTGTACATGG 960

Db  1603  GACCTGCTCGTTCTTCAACAACACGGGCGCCCCGCCCCGGCGCAACTGGTTCGGCGTGTACATGG 1662
               |||||||||||||||||||||||||||||
Qy   961  ATGAATAGCACCGGGGTTCACCAAGAGGTGTGGGGGGGCCCCCCCGTCGCAACATGGGGGGGGTC 1020

Db  1663  ATGAATAGCACCGGGGTTCACCAAGAGCATGTGGGGGGGGCCCCCCCGTTGTAACATCGGGGGGGTC 1722
               |||||||||||||||||||||
Qy  1021  GGCAACAACACTTTGATCTCCCCACGGACTGCTTCCGGGAAGCATCCCGAAGGCCACTTAC 1080

Db  1723  GGCAACAACACCCTGACCTGCTTCCGGGAAGCACCCCGAGGGCTACCTAC 1782
               |||||||||||||
Qy  1081  ACCAAAATGCGGTTCGGGGGGCCTTGGTTG 1107

Db  1783  ACAAAATGTGGTTCGGGGGCCTTGGCTG 1809
```

```
Qy    1  TCCTTCCCGCGTGCTTGGGTAGGTAGGCGGCTCACTCCCAACGCGTCCGCGCGGGACGACCGAGCATCCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db   703  TCCTTCCCGCGTGCTGGGTAGGCGGCTCACTCCCACGCGTCCGCAGCGACCGGCAGGAAGTCACCATCCC 762

Qy   61  ACTGCGACAATACGACGCCCACGTCAGTTCGATTTGCTCGTTCGTTGGGGCGGCTGCCTTCTCCGCT 120
              |||| ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   763  ACCACGACGATACGACGCCCACGTCGATCTCGTCGTTCGTTGGGGCGGCTGCTTTCTGTTCGCT 822

Qy  121  ATGTACGTGGGGGATCTCTCGGGATCTGTTTTTCCTCGTGTTCTGTTCAGCTTCG 180
              |||||||| |||||||||||||||||| ||||||||| |||||||||||||||||
Db   823  ATGTACGTGGGGGAACCTTCTGGGGATCTGTTTTTCCTCGGTCTCTCAGCTTGTTCACCTTCG 882

Qy  181  CCTCGCCCGACATCAGACGGGTACAGGACTGCAATTGTTCAATCTATCCCGCCACGTATCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   883  CCTCGCCCGCCATGTGACATTACAGGACTGTAACTGCTCAATTTATCCCGGCCATGTGTCG 942

Qy  241  GGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAGCAGCCCTAGTGGTA 300
              |||||||||||||||||||| ||||||||||||||||||||||||| ||||||||||||
Db   943  GGTCACCGTATGGCTTGGGACATGATGATGAACTGGTCGCCCACAACAGCCCTAGTGGTG 1002

Qy  301  TCGCAGCTACTCCGGATCCCACACAAGCTGTCCTGGACAATGGTTGGCGCGGGGGGCCCGGGA 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1003  TCGCAGTTACTCCGGATCCCACAAGCCGTCGTCGGACATGGTGGCGGCGGGCCCACTGGGGA 1062

Qy  361  GTCCTGGCGCGGGGCGCCTTGCCTCGGTCTCAGTTGGGGAACTGGGGTAAGGTCTTTGGTGTTGTG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1063  GTCCTGGCGGGGCGCCTTGCCTTACTATTCCATGGCGGGAACTGGGCTAAGGTTCTGATTGTG 1122

Qy  421  ATGCTACTCTTTGCCGGGCGTTGACGGGGAACCTTACACGACGAGGGGGGGACACGGGCCGC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1123  ATGCTACTCTTTTGCTGGGCGTTGACGGGGGATCACCACGTGACCAGGGGGGGGGGGGGGCCAAGCCAA 1182

Qy  481  GCCGGCCCACGGCGCTTACATCCCTCTTCACCACCTGGCCCGGCGTCAGAAAATCCAGCTTGTA 540
              |||| ||||||||||||||||||||||| |||||||||||||| ||
Db  1183  ACCACCAAACAGGCTCGTGTCGTCCATGGCGCAAGTGTGGGGCCTAAGGCTAAGGTCTTTGGTGTTGTG 1242

Qy  541  AACACCCAACGGCGACTGGCACATCAACACGGAACTGCCTTGAACTGCAAATGACTCCCTCCAA 600
              ||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  1243  AACACCAAATGGGAGTTGGCACATCAACACGACTGCCCTGAACTGCAATGACTCTCTCCAG 1302

Qy  601  ACTGGGGTTCCCTTGCCGCGCGTGTTCTACACGGACAGGTTCAATGCGTCGGGATAGCTCAGAG 660
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  1303  ACTGGGGTTTCTTGCCGCGCGCGGCGGTTGCCCTGTTTCTACACATAGTTTCAAACTCGTCGCCCAGAG 1362

Qy  661  CGCATGGCCAGCTGCCGCCCCATTGACCAGTTCGATCAGGGGGTGGGGTCCGAG 720
              ||||||||||| |||||||||||||||||| |||||||||||
Db  1363  CGCATGGCCCAGTGCCGCACCCATTGACAAGTTCGACCAGGGGATGGGGGTCCCATTACTTAT 1422

Qy  721  AATGAGTCCCACGGCTTGGACCAGAGGCCCTATTGCTGGGCACTACGGCACCTCAACCGTGT 780
```

RESULT 13

US-08-904-686A-11

; Sequence 11, Application US/08904686A
; Patent No. 5998130
GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686A
; FILING DATE: 01-AUG-1997
; PRIOR APPLICATION DATA:
; FILING DATE: 18-OCT-1994
; APPLICATION NUMBER: US 08/324,977
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; FILING DATE: 31-AUG-1990
; APPLICATION NUMBER: JP 2-230921
; PRIOR APPLICATION DATA:
; FILING DATE: 09-NOV-1990
; APPLICATION NUMBER: JP 2-305605
; PRIOR APPLICATION DATA:
; FILING DATE: 30-JUL-1993
; APPLICATION NUMBER: US 08/099,706
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6039 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from genomic RNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..6039
; OTHER INFORMATION: /note: "sequence = 333 - 6371 of
; OTHER INFORMATION: SEQ ID NO: 1"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6039

US-08-904-686A-11

Query Match   80.1%; Score 886.2; DB 2; Length 6039;
Best Local Similarity   87.5%; Pred. No. 2.7e-242;
Matches 969; Conservative   0; Mismatches 138; Indels   0; Gaps   0;

```
Qy      1 TCCTCCCCGTGCTGGGTAGGCGTCACTCCCACGCGCTCGCGGGCCAAGGAGCGCCAGCATCCCC 60
          |||||||||| |||||||||||| ||||||||||||||||| || ||||||||||| ||||||||
Db    703 TCCTCCCCGGTTGCTGGGTAGGCCTCACTCCCACGCGCTTCGCAGCGGAAACGTCACCATCCCC 762

Qy     61 ACTGCGACAATACGACGGCCACGTCGATTCGATTCGTTGGGCGGCGCCCTTCGTCCGCT 120
          |||||| |||||||||||||| ||||||||||||||||| |||||||| ||||||| ||||||
Db    763 ACCAGGACGATACGACGCACGTCGATCTCGCTCGTTGTGGCGGCGCTTTCGTCGTTCGGCT 822

Qy    121 ATGTACGTGGGGGGATCTCTGCGGGATCTGTTTTTCTTCGTCTTCTCAGCTGTTCACTTCTCG 180
          ||||||| |||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||
Db    823 ATGTACGTGGGGGACCTCTGCGGGATCTGTTTTTCCTGGTCTTCTCAGCTGTTCACTTCTCG 882

Qy    181 CCTCGCGCGACATCAGACGGGTACAGGAGACTGCAAATTGTTCAATCTATCCGGGCACGTATCA 240
          ||||| |||| |||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db    883 CCTCGCGCGGCATGTGACATTACAGGACACTGTAAACTGCTCAATTTATCCCGGCCCATGTGTCG 942

Qy    241 GGTCACCGCATGGCTTGGGGATTGATGATGATGAAACTGGTTCACCTACAGCAGCCCTAGTTGGTA 300
          |||||||||||||| |||||||| || |||||||| |||||||||||| |||||| |||||||||
Db    943 GGTCACCGTATGGGCTTGGGGACACATGATGAACTGGTTGGCGCCCACAACAGCCCTAGTGGCTG 1002

Qy    301 TCGCAGCTACTCCGGATCCCCACAAGCCGTGTCGTGCAGCATGTGGGAGCCCCCACTGGTGTCG 360
          |||||||||| || |||||||||| |||||||||||| ||||||| |||||| |||||||||||
Db   1003 TCGCAGTTACTCCGGATCCCCACAAGCCGTCGGACATGGTGTGGGGGACCCCCACTGGTGTGTG 1062

Qy    361 GTCCTGGCGGCGCCCTTGCCTACTATTCCATGGTGGGGGAACTGGGCGCTAAGGTCTTGGTTGTG 420
          |||||||||||||||||||| |||||||||||||||||||||| || ||||||||||| |||||
Db   1063 GTCCTCGGGCGGCCCTTGCGCTACTATTCCATGGCGGGGAACTGGGCTAAGGTTCTTGATTGTG 1122

Qy    421 ATGCTACTCTTTGCCCGCGGTTGACGGGACCTTACACGACACAGGGGGGGACACACGCCGCCGC 480
          || ||||||| ||||||||| |||| |||| |||||||||||||| ||||| |||||||| ||
Db   1123 ATGCTACTTTTTGCTCCGGGTTGACGGGGATACCCACGTGACGAGGGGGGGCCAAGCCAAA 1182

Qy    481 GCCGCCCACGGGCCTTACATCCCTCTTCACACCTGGGACGGGAAAATCCAGGTTGTA 540
          |||||| ||||||||||||||||| ||| |||||||||||| || ||||||||| ||
Db   1183 ACCACCAAACAGGCTCGTTGTCCATTGTTCGCAAGTGGGGCCGGTTCAGAAAATCCAGCTTATA 1242

Qy    541 AACACCAAACGCACGGCAGTGTGGCACATCAACAGAGAACTGCCTTGAACTGGCAATGAACTGG 600
          ||||||| ||||| |||||||||||||||||||||| |||||||||| |||||||||||||||||
Db   1243 AACACCAATGGGAGTTGGCACATCAACAGAACTGCCCTGAACTGCAATGACTCTCTCCAG 1302
```

```
Db   1423 GCTGGAGTCTAGCAGATCAGACCAGAGGCCATATTGCTGGCACTACCCACCTCCACAAGTGT 1482
          |||||    |||   |||||||||||| |||   |  |||||||||||||||||||||||||| |||
Qy    781 GGTATCGTGCCCGCGGTTGCAGGTGTGTGCCCAGTGTACTGTTTCACTCCAAGCCCTGTT 840

Db   1483 ACCATCGTACCTGCGCTCGGAGGTGTGCGGGGCCCCAGTGTACTGCTTCCAAGCCCTGTC 1542
          ||||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Qy    841 GTGGTGGGGACGACGATCGTTTCGGCGCCCCTTAGGTACGAGTAGAATGAGACG 900

Db   1543 GTCGTGGGGACGACCGATCGTTTCGGTGTCCCTACGTATAGATGGGGGAGAACGACT 1602
          ||| ||||| |||||||||||||||| |||||||||||||||||||||||||||||||
Qy    901 GACGTGCTGCTTCTCAACAACACGCCGCGCCGCCAGGCGAACTCGGTTCGGCGTACATGG 960

Db   1603 GACGTGCTGCTGCTCAACAACACGGGCCCCGCGCAAGGCAACTGGTTCGGCGCACATGG 1662
          |||||||||||||| |||||||||||| ||| |||||||||||||||| |||| |||||||
Qy    961 ATGAATAGCACCGGGTTCACCAAGACGTGTGGGGGGCCCCGGGGGGGGGTCGGGGGGGTC 1020

Db   1663 ATGAATAGACACCGGGTTCACCAAGACATGTGGGCCCCGGGGGGGGTCCGGGCTGCACATGG 1722
          |||||||||  ||||||||||||||| ||||||  |||||||||||  ||||  |||| |||||
Qy   1021 GGCAACAACACTTTGATCTGCCCCACGGACGACTGCCCGGAGCGAGCCTCCGAGGCCCACTTAC 1080

Db   1723 GGCAACAACACCCTGACCTGCCCCACGGACTGCTTCCGGGAGAGCACCCCCGAGGCTTACCTAC 1782
          |||||||||| ||||| ||||||||||||| || ||||||||| ||||   |||||||||||||
Qy   1081 ACAAAATGTGGTTCGGGGGCGGGCCTTGGCTG 1107

Db   1783 ACAAAATGTGGTTCGGCACATCAACAGAGACTGCGCCCTGAACTGCAATGACTCTCTCCAG 1809
```

RESULT 14
US-09-315-850-11
Sequence 11, Application US/09315850
Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686
; FILING DATE: 01-AUG-1997
; APPLICATION NUMBER: US 08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6039 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from genomic RNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..6039
; OTHER INFORMATION: /note: "sequence = 333 - 6371 of
; OTHER INFORMATION: SEQ ID NO: 1"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6039
US-09-315-850-11

Query Match   80.1%; Score 886.2; DB 3; Length 6039;
Best Local Similarity   87.5%; Pred. No. 2.7e-242;
Matches 969; Conservative   0; Mismatches 138; Indels   0; Gaps   0;

Qy   1 TCCTCCGCTGCTGGGTAGCCGCCTCACTCCCACGCTGCGCGCCAAGGACGCCAGCATCCCC 60
          ||||||| |||||||||||||||| |||||||||||||||| |||| |||||||||||||||
Db   703 TCCTCCGCTGCTGGGTAGCCGCCTCACTCCCACGCTGCGCGCCAAGGACGCCAGCATCCCC 762

Qy   61 ACTGGGACACAATACGACGCGCCACGTCGAATTGCTCGTTGGGCGCTGCCTTCTGCTCCGGT 120
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   763 ACCACGACGATACGACGCCACGTCGATCTCGCTCGTTGGGCGCTGCCTTCTGTTGTCCGGCT 822

Qy   121 ATGTAGGTGGGGGATCTCTGCGGATCTGTTGTTTTCCTCGTGTTCAGCTGTTCACCTTCTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   823 ATGTACGTGGGGGACCTCTGCGGATCTGTTTTTCTTCGTGTTCAGCTGTTCACCTTCTCG 882

Qy   181 CCTCGCCGACATCAGACGTACAGCACGACTGCAATTGTTCAATCTATCCGGCCACGTATCA 240
          |||||| ||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db   883 CCTCGCCGCCATGTGACATTACAGACGACTGTAACTGCTCGAATTTATCCGGCCATGTGTCG 942

Qy   241 GGTCACCGACATGGCTTGGGATATGATGAAACTGGTCACCTACAGCAGCGCTCCGGCGGTA 300
          ||||||||| ||| |||||||||||||||| |||||||||||||| ||||||||||| |||
Db   943 GGTCACGTATGGCTTGGGACATGATGAAACTGGTCGCCACAACAGCCCTAGTGGTG 1002

Qy   301 TCGCAGCTACTCCGGATCCCACAAGCTGTCGTGGACATGGTGGTTGGGGACCTGGGGGA 360
          |||||||||||||||| |||||||||||||||||||||||||||||||||||| |||||
Db   1003 TCGCAGTTACTCCCGGATCCCACAAGCCGTCGTGGACATGGTGGGTTGGGGCCCCACTGGGGA 1062

Qy   361 GTCCTGGCGGCGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCGGCCTAAGGTCTTGGTG 420
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   1063 GTCCTGGGCGGGCCCTTGCCTACTATTCCATGGCGGGGAACTGGGCTAAGGTTCTGGTG 1122

Qy   421 ATGCTACTCTTTGCCGGCGGCGTTTGACGGGGAACCTTACGACGACGAGGGGACACGGCGGC 480

```
Db  1123 ATGCTACTTTTTGCTGGGCGTTGACGGGGGATACCCACGTGACAGAGGGGGGCGCAAGCCAA 1182
             |||||||  ||  |  |||  |||||| ||  ||   |||||||||      ||  |||
Qy   481 GCCCCCCACGGGCTTACATCCCTCTTCACACCTGCCGGCGCTCAGAAAATCCAGCTTGTA 540

Db  1183 ACCACCAACAGGCTCGTCGTTGTCCATGTTCGCAAGTGGGCCGTTCTCAGAAAATCCAGCTTATA 1242
             |||  |||| |||  |    |||| ||||||||||  ||||| |||| | | |||| |
Qy   541 AACACCAACGGGCAGCTGGGCACATCAAACAGAACTGCCTTGAACTGCAATGACTCCCTCCAA 600

Db  1243 AACACCAATGGGAGTTGGGCACATCAAACAGGACTGCCCCTGAACTGCAATGACTCTCTCCAG 1302
             ||||||  ||| ||||||||||||||||||   ||||||||||||||| |||||  ||||
Qy   601 ACTGGGTTCCTTGCCCGCTGTGTTCTTACACCGCCACCAGGTTCAATGCGTCCGGATGGCTCAGAG 660

Db  1303 ACTGGGTTTCTTGTTCGCGGCGCTGTTCTACACACATGTTCAACTCGGTCCGGATGGCTCAGAG 1362
             |||||| | ||| ||||  ||  |||  | |||| || |||||| ||||| ||||||||||||
Qy   661 CGCATGGGCCAGCTCCGCGCCCCCATTGACCAGTTCCATCAGGGGTGGGGTCCCATCACTTAT 720

Db  1363 CGCATGGGCCCAGTGCGCACCATTGAACAAGTTCGACCAGGGATGGGGTCCCATTACTTTAT 1422
             |||||||||| ||| |||||||||| || |||||| ||||||| ||||||||||| |||||
Qy   721 AATGAGTCCCACGGCGTTGACCAGAGGGCCCTTATTGCTGCCACTACGCCACCTTCAACCGTGT 780

Db  1423 GCTGAGTCTCAGACGATCAGACCAGAGGCCCATTATTGCTGGGCACTACCACAAATGT 1482
             ||| ||||  |||  |||  |||||| ||  || | |||| ||||| ||||  |||
Qy   781 GGTATCGTGCCGCGGTTGCAGGTGTGTTGGGCCCAGTGTACTGTTTCACTCCAAGCCCTGTT 840

Db  1483 ACCAATCGTACCTGCCTCGGAGGTGTGCCGCCTCAGTGTACTGCTTCACCCCAAGCCCTGTC 1542
             ||||  || |||||||| ||| |||||| |||| |||||| |  | ||| |||||||| |
Qy   841 GTGCGTGGGGACGACCGATCGTTTCGGCGCGCCCCTACGTACAGAGATGGGGTGAGAATGAGACG 900

Db  1543 GTCGTGGGGACGACGATCGTTTCGGTGTCCCTACGTATAGATGGGGTTCGGCTTGTACATGG 1602
             |||||||||||||| ||||||||| |  ||||||| | |||||||| ||| ||||||| ||
Qy   901 GACCTGCTCGTTCTCAACAACACGCGGCGGGCGCCCCACGGGGCAACTGGTTCGTTGCTGTTACATGG 960

Db  1603 GACCTGCTCGTCAACAACACGCGGCGCCCCCGCAAGGCAACTGGTTCGGTGCTGCACTGG 1662
             |||||||||| |||||||||||||| | || || || ||||||| || |||| ||||||
Qy   961 ATGAATAGCACGGGGTTCACCAAGAGCGTGTGGGGGCCCCCCCGTGTAACATCGGGGGGGTC 1020

Db  1663 ATGAATAGCACCGGGTTCACGCAAGCATGTGGGGGCCCCCCGGGTGTAACATCGGGGGGGTC 1722
             |||||||||| ||||||||  | ||| |||||||||||||| ||||||||||||||||||
Qy  1021 GGCAACAACACTTTGATCTGCCCCACGGACTGCTTCCGGAAGCATCCCGAGGCCCACTTAC 1080

Db  1723 GGCAACAACACCCTGACCTGCGCCCCACGGACTGCTTCCGGAAGCACCCCGAGGCTACTAC 1782
             |||||||||| |||| |||| ||||||||||||||||||||||||| ||||||||| |||
Qy  1081 ACCAAAATGCGGTTCGGGGGCCTTGGTTG 1107

Db  1783 ACAAAAATGTGGTTCGGGGGCCTTGGCTG 1809
```

Query Match   80.1%; Score 886.2; DB 1; Length 9030;
Best Local Similarity 87.5%;
Matches 969; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy     1 TCCTCCCCGCTGCTCGGGGTAGGCCGTCACTCCCACGCTCGCGGCGCCAAGGACGCCAGCATCCCC 60
            |||||  ||||||| || ||| || || | ||| || ||||||||||| ||||||||||||||
Db   703 TCCTCCCCGCTGCTGGGGTAGGCCTCACTCCCCACGCGTTGCTCGTTGGGGGCGGCGCCTTCGCCGCT 762

Qy    61 ACTGGCACAATACGACGCCACGTCGATTTGCTCGTTGGGGGCGGCGCTTTCGTTCGTTCCGGCT 120
            ||||||||| || |||| ||| || ||||||||||||||||||| || || || ||| ||||
Db   763 ACCACGACGATACGACGGCACGTCGATCTGCTGGTTGGGGGCGGCGCTTTCGTTCGTTCCGGCT 822

Qy   121 ATGTACGTTGGGGGGATCTCTCGCGGGATCTGTTTTTCCTCGGTCTCTCAGCGTTTCAGCTTTCTCG 180
            ||| ||| |||||||||  ||  |||||| || || ||| ||||||||| ||| || |||||||||
Db   823 ATGTACGTTGGGGGACCCTCTCGGGGATCGTGTTTTTCCTCGGTCTCTCAGCGTTGTTCAGCTTTCTCG 882

Qy   181 CCTCGCCGCACATCACGAGACGGTACAGGACTGCAAATTGTTCAATCTATCCCGGGCCACGTATCA 240
            ||| ||| || |||| |  || |||| || |||| ||||||||  |||||||| ||||||||| ||
Db   883 CCTCGCCGCGGCATGTGACATTACAGGACTGTAACCTGCTCAATTTATCCCGGCCCCATTGTCG 942

Qy   241 GGTCACCGCATGCCTTGGGGATTATGATGATGAAACTGGTCACCTACACGCAGCCCTAGTGGTA 300
            ||||||  || ||| ||||||||| || ||||| ||| || ||||||||||| |||||||||
Db   943 GGTCACCGTATGGCTTGGGACATGATGAAACTGGTCGCCCACAACAGCCCTAGTGGTG 1002

Qy   301 TCGCAGCTACTCCGGATCCCACACAAGCTGTCGTGGTGGACATGGTGGGCGGGGGGCCCACTGGGGA 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ...
```

```
Db  1003 TCGCAGTTACTCCGGATCCCACAAGCCGTCGTGGACATGGTGGCGGGGGCCCACTGGGGGA 1062

Qy   361 GTCCTGGCGGGGGCCTTGCCTACTATTCCATGGTGGGGGAACTGGGGCTTAAGGTCTTGGTTGTG 420
         ||||||||  ||||
Db  1063 GTCCTGGCGGGGGCCTTGCCTACTATTCCATGGCGGGGGAACTGGGGCTTAAGGTTCTGATTGTG 1122

Qy   421 ATGCTTACTCTTTTGCCGGGGCCTTTGACGGGGGAACCTTACACGACAGGGGACACACGGGCCGC 480
         |||||||  ||
Db  1123 ATGCTACTTTTTGCTGGGCGTTGACGGGGGATACCCACGTGACAGGGGGGGGCAAGCCAAA 1182

Qy   481 GCCGCCCACGGGGCTTACATCCCTCTTCACACCTGGGCGCTCAGAAAATCCAGCTTGTA 540
         |||||||||  |||||
Db  1183 ACCACCACACAGGGCTTCGTCGTTCCATGTTCGCAAGTGGGGCCGTCTCAGAAAAATCCAGCTTATA 1242

Qy   541 AACACCAAGGGCAGCTGGGCACATCAACAGAACTGCCTTGAACTGCAAATGACTCCCTCCAA 600
         ||||||||   ||||||
Db  1243 AACACCAAATGGGGAGTTGGGCACATCAACACGGAACTGCCCTGAACTGCAAATGACTCTCTCCAG 1302

Qy   601 ACTGGGGTTCCTTGCCGCGCTGTGTTCTACACGCACAGGTTCAAATGGTCCGGATGCTTCAGAG 660
         ||||||||  |||
Db  1303 ACTGGGGTTTCTTGCCGCGCTGTTCTACACACATAGTTTCAAACTCGTCCGGGTGCCCAGAG 1362

Qy   661 CGGCATGGGCCAGCTTGCCGCCCATTGACCAGTTCGGATCAGGGGGTGGGGGTTCCCATCACTTAT 720
         |||||||||||  |||
Db  1363 CGGCATGGGCCCAGTGCCGCCACCATTGACAAGTTCGACCAGGATGGGGGTTCCCATTACTTAT 1422

Qy   721 AATGAGTCCCACGGCTTGACCAGAGGCCCTATTGCTGGCACTACGCACCTCAACCGTGT 780
         ||||||||||  ||||
Db  1423 GCTGAGTCTTAGCAGATCAGACCAGAGGCCCATATTGGCTGGCACTACCCACCTTCCACAAATGT 1482

Qy   781 GGTATCGTGCCCCCGCGTTGCAGGTGTGTTGGCCCCAGTGTACTGGTTTCACTCCAAGCCCTGTT 840
         ||||||||  ||
Db  1483 ACCATCGTACCTGGGTCGGAGGTGTGCGGCCCAGTGTACTGCTTCACCCCAAGCCCTGTC 1542

Qy   841 GTGGTTGGGGGACGACCGGATCGTTTCGGCGCCCCCTTACGTACAGATGGGGTGAGAATGAGACG 900
         ||||||||  |||
Db  1543 GTCGTGGGGGACGACGGATCGTTTCGGGTCCCTACGTATAGATGGGGGAGAAGAACGAGACT 1602

Qy   901 GACGTGCTTCTTCTCAACACACGCGGCCGCCCACGGGGCAACTGGTTCGGCTGTTACATGG 960
         ||||||||  ||
Db  1603 GACGTGCTGCTGCTTCAAACAACACCGGCGGCGCGCCCGCAAAGGCAACTGGTTCGGCGTTGCAACATGG 1662

Qy   961 ATGAATAGCACACCGGGTTCACCAAGACACGTGTGTGGGGCGCCCCCCGTGTCAACATCGGCGGGGGGGTC 1020
         ||||||||  ||
Db  1663 ATGAATAGCACCCGGGTTCACCCAAGACATGTGGGGGGGCCCCCCGTGTAACATCGGGGGGGGTC 1722

Qy  1021 GGCAACAACACTTTGATCTGCCCACGGGACTGCTTCCGGAAAGCATCCCGAGGCCCACTTAC 1080
         ||||||||  ||
Db  1723 GGGAACAACACCCTGACCTGCCCCACGGGACTGCTTCCGGAAGCACCCCGGAGGCTACCTAC 1782

Qy  1081 ACCAAAATGCGGTTCGGGGGCCTTGGTTG 1107
         ||||||||  ||||||||||||
Db  1783 ACAAAAATGTGGTTCGGGGGCCTTGGCTG 1809
```

Search completed: February 21, 2005, 09:05:30
Job time : 231 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 05:19:09 ; Search time 4250 seconds
                                             (without alignments)
                                             9914.623 Million cell updates/sec

Title: US-09-664-363-19
Perfect score: 1107
Sequence: 1 TCCTCCCGCTGCTGGGTAGC............GCGGTTCGGGGGGGCCTTGGTTG 1107
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_htc:*
               4: gb_est3:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_gss1:*
               9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	% DB	ID	Description
1		137	12.4	488	1	AV755731	AV755731 AV755731
2		104.4	9.4	492	1	AV758366	AV758366 AV758366
3		58	5.2	179	7	R28798	R28798 F0-266D 22
4	c	45.2	4.1	839	9	CNS004NB	AL054280 Drosophil
5	c	42.4	3.8	925	9	CNS091P	AL053013 Drosophil
6	c	42.2	3.8	731	6	CB651667	CB651667 OSJNEb16L
7	c	41.4	3.7	864	9	CNS01219	AL173718 Tetraodon
8	c	40.8	3.7	628	8	BZ363741	BZ363741 id90b10.g
9	c	40.8	3.7	642	3	CR721023	CR721023 Tetraodon
10	c	40.8	3.7	774	9	BZ797506	BZ797506 PUFJQ7BTB
11	c	40.8	3.7	837	8	BZ636670	BZ636670 PUFJQ7BTB
12	c	40.8	3.7	855	9	CC692345	CC692345 OGCCW77TC
13	c	40.8	3.7	867	9	CC692352	CC692352 OGMBB44TH
14	c	40.8	3.7	897	9	CC977474	CC977474 OG3BB55TH
15	c	40.8	3.7	914	9	CG286439	CG286439 ZUAB789TH
16	c	40.4	3.6	409	5	BQ469708	BQ469708 H20IK01r
17	c	40.4	3.6	409	5	BQ469721	BQ469721 H20IK19r
18	c	40.4	3.6	675	7	CN126631	CN126631 RHOH1_18
19	c	40.4	3.6	720	5	BQ578584	BQ578584 WHEO307_A
20	c	40.4	3.6	746	9	CL182015	CL182015 104_393_1
21	c	40.4	3.6	752	7	CK124990	CK124990 BES182411
22	c	40.4	3.6	1100	9	CNS016KD	AL106855 Drosophil
23	c	40.2	3.6	803	6	CA212082	CA212082 SCSFAD110
24	c	39.8	3.6	667	6	CA293368	CA293368 SCAGFL804
25	c	39.8	3.6	997	9	CNS006DN	AL065132 Drosophil
26	c	39.4	3.6	688	9	CA752491	CA752491 UI-M-FO0-
27	c	39.2	3.5	925	9	CNS009lP	AL053013 Drosophil
28	c	38.8	3.5	375	1	AA408413	AA408413 EST02883
29	c	38.8	3.5	801	9	CC635330	CC635330 OGUIL59TV
30	c	38.8	3.5	830	8	CC386543	CC386543 PUFMR257B
31	c	38.8	3.5	851	9	CG080859	CG080859 PUFMR70TD
32	c	38.8	3.5	872	9	CG451056	CG451056 OG9AA89TH
33	c	38.8	3.5	888	9	CG257405	CG257405 OGYCD567V
34	c	38.8	3.5	926	9	CG115481	CG115481 PUJCE95TB
35	c	38.8	3.5	955	9	CG451068	CG451068 OG9AA89TV
36	c	38.8	3.5	964	9	CG231178	CG231178 OGYCD45TV
37	c	38.6	3.5	994	9	CG080858	CG080858 PUFMR70TB
38	c	38.6	3.5	512	6	CA664299	CA664299 w1k1.pk00
39	c	38.6	3.5	617	9	CL169972	CL169972 104_370_1
40	c	38.6	3.5	634	7	CN139610	CN139610 OX1_23_C0
41	c	38.6	3.5	743	7	CN139693	CN139693 OX1_23_C0
42	c	38.4	3.5	655	9	CG274746	CG274746 OGWAU3BTV
43	c	38.4	3.5	760	7	CK470808	CK470808 AGENCOURT
44	c	38.4	3.5	782	7	CC605275	CC605275 OGUHI31TH
45	c	38.4	3.5	860	9	CG286130	CG286130 OGXEPI6TV

ALIGNMENTS

RESULT 1
AV755731/c
LOCUS   AV755731   488 bp   mRNA   linear   EST 19-OCT-2000
DEFINITION AV755731 BM Homo sapiens cDNA clone BMFAKB03 5'', mRNA sequence.
ACCESSION   AV755731
VERSION   AV755731.1 GI:10913579
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 488)
AUTHORS   Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
            Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Zhou,M.,
            Zeng,L., Xu,S., Gu,W., Tu,F., Jia,J., Fu,G., Ren,S., Zhong,M.,
            Lu,G., Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and
            Chen,Z.
TITLE   Homo sapiens cDNA BM clones
JOURNAL   Unpublished (2000)
COMMENT   Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-5080919(ex.45)
            Fax: 86-21-5080922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES   Location/Qualifiers
     source   1..488
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BMFAKB03"
                     /tissue_type="Bone marrow"
                     /cell_type="CD34+ hematopoietic stem/progenitor cell"
                     /lab_host="BM25.8"
                     /clone_lib="BM"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
ORIGIN

Query Match   12.4%; Score 137; DB 1; Length 488;
Best Local Similarity   60.0%;
Matches 280; Conservative 0; Mismatches 182; Indels 5; Gaps 3;

Qy   94 GTTGGGGGCGGCTGCCTTCTGCTCCGCTATGTACGTGGGGGGATTCTGCGGATCGTGTTTC 153
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||| |||||||
Db   472 GTGGTVGTCACACTCGCTCTGCTCTGCTCAGCTCTCTACGTGTGTGGGACCTCTGCGACGGGAGTGATG 413

RESULT 2

```
Qy   154  CTCGTCTCTCAGCTGTGTTCACCTTCTCGCCTCGCCGACATCAGACGGTACTAGGACTGCAAT  213
          ||||||||||||||| |||   ||||||||||||||||||||||||||||||||||||||||||
Db   412  CTTCCAGTTCAGCTGATCA---TCTGGCCTCAGCAGTTGTGCATGAGTGATGAATGCAAC    356

Qy   214  TGTTCAATCTATCCCGGCCACGTATACAGGTCACCCGCATG-GCTTGGGATATGATGATGAA  272
          ||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||||
Db   355  TGCTCCAATCTATCCTGGCCCATCAGACACCGTATACAGCATGGGACACTGAGATGATGAA   296

Qy   273  CTGGTCACCTACACAGCAGCCCTAGTGTATCGTCAGTTACTCCGGATCCCACAAGTGTCGT  332
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   295  CTGGTCCTGCTGCACCCGCTCGTATGATCATGGGCTACCGTTCCTCTGAGGTCATCAT     236

Qy   333  GGACATGGTGGCGGGGGGGGGCTCTGGGGGGGGCCCAGAGTCGGGGGGCCTTACTACTCATGGT  392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   295  AGATATCATCAGCGGGGGCCTCACTCGGGCGTCATGTTCGGCTTTATCTCATTCTTATGCA  176

Qy   393  GGGGAACTGGGCTAAGGTCTTGGTTGTTGGATGTCTACTCTTTGCCGGCGTTGACGGGGGAACC  452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   175  GGGGAGCGTGGGAAAGTCGGCTGTCATCCTTCTGTTCACCGTCAACGGCAGCAC       116

Qy   453  TTACACGACAGGGGGGGGAC-ACACGCGCGCCGCCGGCTTACATCCCTCTTCTTCACAC      511
          |||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db   115  CCACCTCACAGGGGTCGGCCACTGCTAGCCGCGCGCGGCCAGTTTCGTTCGCTTCAGCTTGCTCACC   56

Qy   512  CTGGGGCCCGGCTCAGAAAATCCAGCTTGTAAAACACCAACGGCCAGCTGG  558
          |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   55   CTGGCGCTAAAACAGAACATCCAACTCATTANGATCTNATTCAATGGG   9
```

LOCUS   AV758366   492 bp   mRNA   linear   EST 19-OCT-2000
DEFINITION   AV758366 BM Homo sapiens cDNA clone BMFAKA03 5', mRNA sequence.
ACCESSION   AV758366
VERSION   AV758366.1 GI:10916214
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 492)
  AUTHORS   Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,Y., Xiao,H.,
             Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
             Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M.,
             Lu,G., Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and
             Chen,Z.
  TITLE   Homo sapiens cDNA BM clones
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Zeguang Han
             Chinese National Human Genome Center at Shanghai
             351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
             201203, P. R. China
             Tel: 86-21-50801919(ex.45)
             Fax: 86-21-50801922
             Email: hanzg@chgc.sh.cn
             This clone is available at CHGC in Shanghai.
FEATURES   Location/Qualifiers
     source   1..492
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BMFAKA03"
                     /tissue_type="Bone marrow"
                     /cell_type="CD34+ hematopoietic stem/progenitor cell"
                     /lab_host="BM25.8"
                     /clone_lib="BM"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
ORIGIN

Query Match   9.4%; Score 104.4; DB 1; Length 492;
Best Local Similarity   56.1%; Pred. No. 1.2e-17;
Matches 254; Conservative   0; Mismatches 194; Indels   5; Gaps   3;

RESULT 3

```
Qy   110  TCTGCTNCCGCTATGTNACGTGGGGGANCTCTGNTGCGAGANCTCTGTTTTCCTGTCTCTCAGCTGT  169
          ||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   457  TGTGANCAGCTNCACTNACGTGTTGGACCTCTGCCGGTTGGGGTATCGGCTTGCAGCCCAACTGA  398

Qy   170  TCACCTTCTCGCCTCGCCGACATCAGACGGTACAGGACGGTACAGGACTGCAATTGTTCAATCTATCCCG  229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   397  TTA---TCTCTCAGCAGCAACATTGGTTCTGTGCAAGAANTGGCTACTCATTCTATCCTG    341

Qy   230  GCCACGTATCAGGTC-ACGCCATGGCTTGGGGATATGATGATGAACTGGTCACCTACACGA  288
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   340  TCCATGATCACTGGACTACAGTAGGCTGGCCAATGGGATCCCATAGAGTCCTCATATGAGTCCTCAGGT  281

Qy   289  GCCCTAGTGTATCGCAGCTACTCCGGATCCCACAAGCTGTCGTCGTGGTGACATGGTGGGGG  348
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   280  TCCATGATACGTGGCGTTGGGGCCGGGAAATGGGCGGTTCCTCATGAACTGGTCGCCGGGGG  221

Qy   349  GCCCCACTGGGGGAGTGCTCCTGGGGGGGGCCTTGCCTACTATTTCCATTGGTGTGGGGGGTAG  408
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   220  GCACACTGCGGGTCATGTTCGGGCTCAGCTTACTTCACAGAGGAGCGGTTGGCCAA  161

Qy   409  GTCTTGGTTGTGATGCTACTCTTTGCCGGCGGGGGTTGACGGGGGAACCTTACACGACAGGGGGG  468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   160  GTAGCTGTCATCCATCTGTTCACCGTCAACGGCAGCAACCACGTNATCAGGTGGT  101

Qy   469  ACACACGCGCGCCGCGCCCCACGGGGGCT-TACATCCCTCTTCACCCTGGGGGCCGGGGCTCAGAA  527
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   100  ATCACTGCGGGCGCGACGACAGCAGTTTCGTCGTCAGGTTGCATAATGCCTGGGGCGGGCTAAGCAGTA  41

Qy   528  AATCCAGCTTGTAAAACACCAACGGCCAGCTGGCA  560
          |||||||||||||||||||||||||||||||||| |
Db   40   CATCCAGCTCATGANCGTCAATGGCAATNGGNA   8
```

LOCUS   R28798   179 bp   mRNA   linear   EST 03-DEC-1999
DEFINITION   F0-266D 22 week old human fetal liver cDNA library Homo sapiens
             cDNA clone F0-266D 5', mRNA sequence.
ACCESSION   R28798
VERSION   R28798.1 GI:6514170
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 179)
  AUTHORS   Choi,S.S., Yun,J.W., Choi,E.K., Cho,Y.G., Sung,Y.C. and Shin,H.S.
  TITLE   Construction of a gene expression profile of a human fetal liver by
             single-pass cDNA sequencing
  JOURNAL   Manm. Genome 6 (9), 653-657 (1995)
  MEDLINE   96081342
  PUBMED   8535075
COMMENT   Contact: Hee-Sup Shin
             Developmental Genetics
             Pohang Institute of Science & Technology
             San31, Hyojadong Pohang, 790-784 Republic of Korea
             Tel: 562-279-2291
             Fax: 562-279-2199
             Email: shinhs@vision.postech.ac.kr
             Seq primer: T3 primer.
FEATURES   Location/Qualifiers
     source   1..179
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="F0-266D"
                     /lab_host="XL1-blue MRF'"
                     /clone_lib="22 week old human fetal liver cDNA library"
                     /note="Vector: pBluescriptII SK(-); Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library made by oligo-dT primed and
                     directionally cloned between 5' ExoR I-XhoI3' sites."
ORIGIN

RESULT 4

```
Qy  521  CTCAGAAAATCCACCTTGTAAACACCAACGGCAGCTGGGCACATCAACAGAACTGCCTTGA  580
         |||||| ||||||| |||| || ||||||||||| ||||||||||||| ||||||||||||
Db   63  CGGAGAAACATTCAGCTCGTRAAACCCTAATGGCAGCTGGCACATCAACAGGACTGCCTTGA  122

Qy  581  ACTGCAAATGACTCC  594
         ||| ||||| |||||
Db  123  ATTGCAAATGAGTCC  136
```

RESULT 4
```
LOCUS       CNS004NB        839 bp    DNA    linear    GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR10B16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL054280
VERSION     AL054280.1  GI:4931788
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 839)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191.91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..839
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR10B16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
Query Match    4.1%;  Score 45.2;  DB 9;  Length 839;
Best Local Similarity  19.5%;  Pred. No. 0.36;
Matches  72;  Conservative  133;  Mismatches  165;  Indels  0;  Gaps  0;

Qy  629  CGCACAGGTTCAATGCGTCCGGATGCTCAGAGCGCGCCATGGCCGCCCCCCCATTGACC  688
         :|:::|:::|:|                    ::::| :|:|
Db  836  CBBTCGSSSTPBBSCSSCCTSSSSSSTSSTSSKTGKGCSKSGSGSBSSGSSSSSSGCSTYSSST  777

Qy  689  AGTTCGATCAGGGGTTGGGGGGGTCCCATCACTTATATAAGTGAGTCCCACGGCGTTGGACAGAGGC  748
         ::|::::|::::::||:::::::|
Db  776  GGGSGSVSGGSGSSGSGGSBBSSGCSSSSSBSSGCSSSSSSSSSSSSGGGGCSTS  717

Qy  749  CCTATTGCTGGCACTCACGCACCTCAACCGTGTGGTATTCGTGCCCCGGTTGCAGGTGTGTG  808
         :|:::::|::::          :::|::::
Db  749
```

```
Db  716  CSSTTCSSBPSBGKSSSGCSGSCSSSTSSTGBSTGYTYTTCTSSSSCSTCSSSGTTCBBSTT  657

Qy  809  GCCCAGTGTACTGTGTTCACTCCAAGCCCTGYTTGTGTGGGGTGGGGGDACGACCGATCGTTTCGGCG  868
         :|:::|
Db  656  SSTSSSTTSTTTSBTYCSCTTCKCCKCCYTCSYTBTTTTSBCBCSCGSCGSGBTTBSCSSSSS  597

Qy  869  CCCTTACGTACAGATGGGGGTGGAGAATGAGAGACGGAGCGTGCTTCTTCCAACACACGGGC  928
         :::|:|
Db  596  SGSTTSSSSSSSSGGSKGGCSTTSSSSSSSTSSSSSGSTSSSSCSTSCBTCCCCSCYSCCCCCSCCS  537

Qy  929  CGCCACGGGGCGGAACTGGTTCGGCTGGTTACATGGATGAAATAGCACCGGGTTCACCAAGACGT  988
         |:::::
Db  536  CTCCCCSYSCSTCCSCCCTSSSSSCTCGSSCSTSSSSSGGSSGTTBGSSTSGCCSS  477

Qy  989  GTGGGGGCCC  998
         :|:::
Db  476  TSTSGTSSSS  467
```

RESULT 5
```
LOCUS       CNS0091P        925 bp    DNA    linear    GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the library
            P1 and EST libraries. A more detailed description of the library,
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
Query Match    3.8%;  Score 42.4;  DB 9;  Length 925;
Best Local Similarity  13.2%;  Pred. No. 2.2;
Matches  44;  Conservative  149;  Mismatches  141;  Indels  0;  Gaps  0;

Qy  226  CCCGGCCACGTATCAGGTCACCGCATGGCTTGGGATATGATCATGAACTGGTCACCTACA  285
         :|:::||:|:|
Db  580  CSSSSSCSCCBCCCCCCSYCCSSSTSBSCSCSSTSBSGCSSCCCCSSVCGTSCSSSSSCCSS  639

Qy  286  GCAGCCCTAGTGGTATCGGCGTACTCCGGATCCCAACAAGCTGTGTCGTGGACATGGTGCG  345
```

RESULT 6

```
LOCUS       CB651667/c
DEFINITION  OSJNEb16L12.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb16L12 3', mRNA sequence.
ACCESSION   CB651667
VERSION     CB651667.1  GI:29646660
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 9595
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR Primers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 16   row: L column: 12
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..731
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb16L12"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     8601)"
ORIGIN
```

Query Match   3.8%; Score 42.2; DB 6; Length 731;
Best Local Similarity   55.9%; Pred. No. 2.4;
Matches   80; Conservative   0; Mismatches   63; Indels   0; Gaps   0;

```
Qy     943 TGGTTCGGCTGTACATGGATGAAATAGCACCCGGTTCACCAAGACTGTGTGGGGCCCCG 1002
Db     640 SSSSTSTSSSSSSTSKSSSSGSSS...
Qy     346 GGGGCCCACTGGGGAGTTCCTGGGCGGGCCCTTGCCTACTTA...
Db     700 SSSSTSTSSSSSVSSSGSSKSSTBSSGSBSSGSS...
Qy     406 AAGGTCTTGGTTGTGATGCTACTCTTGCCGGGGTTGACGGGGAACCTTA...
Db     760 CCTCCSYSYSSSSSTSSSSSTSWGSTSGSSSSSVG...
Qy     466 GGGACACACGGCGGCCGCGCCGCGGGCCTTACATCCTCTTCACACTTG...
Db     820 MBCYTSTSCCGSSSSSGSKGGVTKCGCGGCGGSSSTNGMBGTSSA...
Qy     526 AAAATCCAGCTTGTAAAACACCAACGGCAGCTGGC 559
Db     880 KSSRASSSVSSSGSSGVSSNSSSASKSSSGSVS 913
```

```
Db     564 TCGTTCGGTCCGCTTCGGGGACAGCAGCGAGCGAGCG...
Qy    1003 TGCAACATCGGGGGGGGTCGGGCCAACAAACACTTTGA...
Db     504 TGGAACATCGGAGGAGGTCGACTTCAACATGTCG...
Qy    1063 CATCCCGAGCGCCACTTACACCAA 1085
Db     444 CCCAGCGACCTCAAGGATAGCAA 422
```

RESULT 7

```
LOCUS       CNS01Z19
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            219D07 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   CNS01Z19
VERSION     AL173718.1  GI:7811775
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontoidei; Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 864)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..864
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="219D07"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG219CB04SP1-end :
                     PUC-Ori"
ORIGIN
```

Query Match   3.7%; Score 41.4; DB 9; Length 864;
Best Local Similarity   42.3%; Pred. No. 4.1;
Matches 150; Conservative   12; Mismatches 193; Indels   0; Gaps   0;

```
Qy     406 AAGGTCTTGGTTGTGATGCTACTCTTGCCGGGGTTGACGGGGAACCTTACACGACAGGG 465
Db      88 ACGCTCTCTGGACGTGGAGGAGTACCTTCATCGACCACGACGAGGAGCMGCCCTACGAGGTGCGT 147
```

RESULT 8

```
Qy    466 GGGGACACACGGGCCCCGCCCCCACCCGGCTTACATCCCTCTTCACCTGGGGGCCGGCTCAG 525
          ||:|: |||||||||||||||||||||||||:|||| ||::|||||| ||||||||||||
Db    148 GACCGACCCCCSCCGGCGCGCCTCCGGCGGCCGCTCCMCTTGTTCCCTCGCTT 207

Qy    526 AAAATCACGCTTGTAAACACCAAACGGGACGCAGCTGGGCACATCAAACAGAACTGCCTTGAACTGC 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    208 GTTTTTCAGAGTGTTTTCCTAAAGAGCGTAGATGCTCTACMACACGGGACGACTTCMGCCAC 267

Qy    586 AATGACTCCCTCCAAAACTGGGTTCCTTGCGCGCGTGTTCTTACACGGCACAGGTTCAAATGCG 645
          |||| ||||||||||||||||| ||||||| || |||
Db    268 AGTGCCCCGGSAGATGGAGGGRGGCCCTGGCCGMGTATTTTGACGGGTACMAGGGGCTGCGGTG 327

Qy    646 TCCGGGATGCTCAGAGGGCGCATGGCCAGCTGCGGCCCCATTGACCAGTTCGGATCAGGGGTGG 705
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    328 GCCGGGATCGGAGGGGCCTCCTACGAGATCCTTCFAGAGTTCMMAGACTTCTWTCCCMCMCCGTGGCM 387

Qy    706 GGTCCCATCACTTATATAATGAGTCCCACGGGCTTGGACCAGAGAGTTCMMAGACTTCTWTCCTGGC 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    388 GGTAAGAGGAGGCCMGATCCACACCGGCCCTGGGGGCCCAGAGAGGGCCCCTTCGCTGTC 442
```

Query Match   3.7%;   Score 40.8;   DB 8;   Length 628;
Best Local Similarity   53.0%;   Pred. No. 5.6;

Matches   87;   Conservative   0;   Mismatches   77;   Indels   0;   Gaps   0;

```
Qy    331 GTGGACATGGTWGGCGCGGGGGGGGCCCACTGGGGGAGTCCTGGGCGGCTTGCCTACTATTCCATG 390
          || || |||||||||||||||||||||||||||||||||||||||||||||
Db    510 GTACCGGTTGTCCCCTGCACTCCACGGGAGCGGGGCGGGGGGGGGCGGCCTTGGCGCTTGGCGAAGAAGCCCGG 451

Qy    391 GTGGGGAACTGGGGCCTRAGGTCTTGGTTGTGTTGTGATGGTGACTACTTCTTTGCCGCGTTGACGCGGGA 450
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    450 GTGGCCGACGTGCGTCACGCACCAGCATCGACCGCGCTTCGCGCGGCGGTAGCCCCTGGAA 391

Qy    451 CCTTACGACGGAGGGGGGGACACACGGCGGGCGCCACGGGGCT 494
          |||||||||||||||||||||||||||||||||||||||||
Db    390 GATGCTGCGCCGCCCGCGCGCCCGGCACGGCGCCGGAGAAGGGCT 347
```

Query Match   3.7%;   Score 40.8;   DB 3;   Length 642;
Best Local Similarity   52.3%;   Pred. No. 5.6;
Matches   90;   Conservative   0;   Mismatches   82;   Indels   0;   Gaps   0;

```
Qy    273 CTGGTCACCTACAGCAGCCCTAGTGGTATGGTATGGCCAGTAACTCCGGATCCCACAAGCTGTCGT 332
          |||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    207 CTGAGCACGRAAGGGAAGCCCTACTGCCCACAAAGCCCTGCTCACGCACGCCTCCAGTCTCGTTGC 266

Qy    333 GGACAATGGTGGCGGCGGGGACCCACTGGGGAGTCCTGGCGGGGCCTTGCCTACTATTCCATGGGT 392
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 TGAGGAGCCGTGCCTGGTGTCACTGCGGGGAGGGTTCTTAACTCCTACTGGGTGGG 326

Qy    393 GGGGGAACTGGGGCTRAGGTCTTGGTTGTGTTGTGATGCTACTCTTTGCCGCGTTGAC 444
          |||||||||||||||||||||||||||||||||||||||||||
Db    327 CGAAGACTCCAACCTACAAGTTTTTCCAGGTGATGATCGACCCCTTCCAC 378
```

```
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 774)
AUTHORS     Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J., and
            Bennetzen,J.
TITLE       Maize Genomics Consortium
JOURNAL     Unpublished (2003)
COMMENT     Other_GSSs: PUFUQ78TD
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Contact: Cathy Whitelaw
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..774
                     /organism="zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa358N11"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

Query Match      3.7%;  Score 40.8;  DB 8;  Length 774;
Best Local Similarity  53.0%;  Pred. No. 5.9;
Matches   87;  Conservative   0;  Mismatches   77;  Indels   0;  Gaps   0;

Qy  331 GTGGACATGGTGGCGCGGGCCCACTGGGGAGTCCTTGGCGGCGCCTTGCCTACTATTCCATG 390
Db  425 GTACCCGTTGTCCCTGCACTCCACGGACGATCGCAGCCCAGCCCGTTGCTTGGCGGAGAAGCCCCG 484

Qy  391 GTGGGGGACTGGGCGTAAGGTCTTTGGTTGTGATGCTACTCTTTTGCCGGGCGTTGACGGGGGAA 450
Db  485 GTGGCCGGAGTCGTCACGACCAGCATCGACCGGCCTCTGCGCCCGGTTGACCCCTGGAA 544

Qy  451 CCTTACACGACCAGGGGGGGGACACACGGCCCGGCGCCCCGCGCCCACGGGGCT 494
Db  545 GATGCTGCCCGCGCGCGCCCGCACGGGCCCGCACGGGCCCGAGAACGGGCT 588


RESULT 11
LOCUS       B2638670     837 bp   DNA   linear   GSS 29-JAN-2003
DEFINITION  OGCCW77TC ZM_0.7_1.5_KB zea mays genomic clone 2MMBMa0143N09,
            genomic survey sequence.
ACCESSION   B2638670
VERSION     B2638670.1  GI:28091427
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 837)
AUTHORS     Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohling,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE       Maize Genomics Consortium
JOURNAL     Unpublished (2002)
COMMENT     Consortium for Maize Genomics
            Other_GSSs: OGMBB44TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Contact: Cathy Whitelaw
            Seq primer: TF
            Class: sheared ends.
```

```
FEATURES             Location/Qualifiers
     source          1..837
                     /organism="zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0143N09"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Query Match      3.7%;  Score 40.8;  DB 8;  Length 837;
Best Local Similarity  53.0%;  Pred. No. 6;
Matches   87;  Conservative   0;  Mismatches   77;  Indels   0;  Gaps   0;

Qy  331 GTGGACATGGTGGCGCGGGCCCACTGGGGAGTCCTTGGCGGCGCCTTGCCTACTATTCCATG 390
Db   14 GTACCCGTTGTCCCTGCACTCCACGGACGATCGCAGCCCAGCCCGTTGCTTGGCGGAGAAGCCCCG 73

Qy  391 GTGGGGGACTGGGCGTAAGGTCTTTGGTTGTGATGCTACTCTTTTGCCGGGCGTTGACGGGGGAA 450
Db   74 GTGGCCGGACCGTCGTCACGACCAGCATCGACCGGCCTCTGCGCCCGGTTGACCCCTGGAA 133

Qy  451 CCTTACACGACCAGGGGGGGGACACACGGCCCGGCGCCCCGCGCCCACGGGGCT 494
Db  134 GATGCTGCCCGCGCGCGCCCGCACGGGCCCGCACGGGCCCGAGAACGGGCT 177


RESULT 12
CC692345/c
LOCUS       CC692345     855 bp   DNA   linear   GSS 19-JUN-2003
DEFINITION  OGMBB44TH ZM_0.7_1.5_KB zea mays genomic clone 2MMBMa0312H16,
            genomic survey sequence.
ACCESSION   CC692345
VERSION     CC692345.1  GI:32097121
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 855)
AUTHORS     Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohling,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE       Maize Genomics Consortium
JOURNAL     Unpublished (2002)
COMMENT     Consortium for Maize Genomics
            Other_GSSs: OGMBB44TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..855
                     /organism="zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0312H16"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Query Match      3.7%;  Score 40.8;  DB 9;  Length 855;
Best Local Similarity  53.0%;  Pred. No. 6.1;
Matches   87;  Conservative   0;  Mismatches   77;  Indels   0;  Gaps   0;
```

```
Qy   331  GTGGGACATGGTGGGCGGGGGGGGGCCCACTGGGGGAGTCCTTGGCGGGGGAGTCCTTGCTACTATTCCATG  390
               |||| | ||||||||||||||| ||||||||||||| ||||| |||| | ||||||
Db   852  GTACCCGGTTGTCCCTGCACTCCACGGGCAGGCCCAGCGCTGCTTGGCGGAAGAGCCCGG  793

Qy   391  GTGGGGAACTGGGGCTAAGGTCTTGGTTGTTGTGATGCTACTCAGCCCGTTGACGGGGAA  450
               ||||||||| |||||||||||||||||||||||||||||||
Db   792  GTGGCCGGACGTCGTCACGCACCACCAGCATCGACCGCCTCGCGCGCCGTAGCCCTGGAA  733

Qy   451  CCTTACACGACGAGGGGGGGACACACGGCCGCGGCCGCCCACGGGGCT  494
               ||||||||||||||||||||||||||||||||||||||||||||
Db   732  GATGCTGCGCGCGCCGCGCCGCGCCGCGCCCGCCGCCCGCCGAGAAGGGCT  689
```

RESULT 13
LOCUS   CC692352   867 bp   DNA   linear   GSS 19-JUN-2003
DEFINITION OGMBB44TV ZM_0.7_1.5_KB zea mays genomic clone ZMMBMa0332H16,
            genomic survey sequence.
ACCESSION   CC692352
VERSION   CC692352.1 GI:32097128
KEYWORDS   GSS.
SOURCE   Zea mays
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 867)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohling,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.,
  TITLE   Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT   Other GSSs: OGMBB44TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..867
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0332H16"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Query Match   3.7%; Score 40.8; DB 9; Length 867;
  Best Local Similarity   53.0%; Pred. No. 6.1;
  Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy   331  GTGGACATGGTGGGCGGGGGGGCCCACTGGGGGAGTCCTTGGCGGGGGAGTCCTTGCTACTATTCCATG  390
Db    16  GTACCCGGTTGTCCCTGCACTCCACGGGCAGGCCCTTGGCGGAAGAGCCCCG  75

Qy   391  GTGGGGAACTGGGCTTAAGGTCTTGGTTGTGATGCTACCCAGCCCGTTGACGGGGAA  450
Db    76  GTGGGCGGACGTCGTCACGCACCACCAGCATCGACCGCCTCGCGCGCCGTAGCCCTGGAA  135

Qy   451  CCTTACACGACGAGGGGGGGACACACGGCCGCGGCCGCCCACGGGGCT  494
Db   136  GATCCTGCGCGCGCCGCGCCCGCGCCGCGCCCGCCGAGAAGGGCT  179
```

RESULT 14
LOCUS   CC977474   897 bp   DNA   linear   GSS 18-AUG-2003
DEFINITION ZUAB7899TH ZM_3.0_4.0_KB Zea mays genomic clone ZMMBPa0012P10,
            genomic survey sequence.
ACCESSION   CC977474
VERSION   CC977474.1 GI:33837352
KEYWORDS   GSS.
SOURCE   Zea mays
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 897)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohling,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE   Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT   Other GSSs: ZUAB7899TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..897
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBPa0012P10"
                     /clone_lib="ZM_3.0_4.0_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 3-4 kb 'unfiltered'
                     genomic DNA library"
ORIGIN

Query Match   3.7%; Score 40.8; DB 9; Length 897;
  Best Local Similarity   53.0%; Pred. No. 6.1;
  Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy   331  GTGGACATGGTGGGCGGGGGGGCCCACTGGGGGAGTCCTTGGCGGGGGAGTCCTTGCTACTATTCCATG  390
Db   658  GTACCCGGTTGTCCCTGCACTCCACGGGCAGGCCCTTGGCGGAAGAGCCCCG  717

Qy   391  GTGGGGAACTGGGCTAAGGTCTTGGTTGTGATGCTACCCACTCTTTGCCGGGGTTGACGGGGAA  450
Db   718  GTGGGCGGACGTCGTCACGCACCACCAGCATCGACCGCCTCGCGCGCCGTAGCCCTGGAA  777

Qy   451  CCTTACACGACGAGGGGGGGACACACGGCCGCGGCCGCCCACGGGGCT  494
Db   778  GATGCTGCGCGCGCCGCGCCCGCGCCGCGCCCGCCGAGAAGGGCT  821
```

RESULT 15
LOCUS   CG286439   914 bp   DNA   linear   GSS 25-AUG-2003
DEFINITION OG3BE55TH ZM_0.7_1.5_KB zea mays genomic clone ZMMBMa0767J13,
            genomic survey sequence.
ACCESSION   CG286439
VERSION   CG286439.1 GI:34200653
KEYWORDS   GSS.
SOURCE   Zea mays
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 914)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohling,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE   Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)

COMMENT   Other GSS8: OG3BE5STV
           Contact: Cathy Whitelaw
           TIGR
           9712 Medical Center Drive, Rockville, MD 20850, USA
           Tel: 301-838-5843
           Fax: 301-838-0208
           Email: whitelaw@tigr.org
           Seq primer: TR
           Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..914
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0767J13"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBC5K-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"

ORIGIN

     Query Match   3.7%; Score 40.8; DB 9; Length 914;
     Best Local Similarity 53.0%; Pred. No. 6.2;
     Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy   331 GTGGACATGGTGGCGGGGGCCCCACTGGGGAGTCCTGGCGGGCCTTGCCTACTATTCCATG 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   116 GTACCCGTTGTCCCTGCACTCCACGGGCAGGCCCAGCCCGCTGCTTGGCGAAAGAAGCCCCGG 175

Qy   391 GTGGGGAACTGGGGCTAAGGTCTTGGTTGTGAATGCTTGTTACTCTTTGCCGGCGTTGACGGGGGAA 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176 GTGGCGGGACGTCGTCACGCACCACGCATCGACCGGCTTCGCGCGCCGCCGCCCGCTAGCCCTGGAA 235

Qy   451 CCTTACACGACGAGGGGGGACACACGGGCCCGCGCCCACCGGGGCT 494
          |||||||||||||||||||||||||||||||||||||||||
Db   236 GATGCTGCCCCGCCCGCCGGCCCCACGGCCCGCGCCGCCGCCAGAAGGGGCT 279

Search completed: February 21, 2005, 09:01:39
Job time : 4258 secs

STIC Search Report
Biotech-Chem Library

TO: Bao-Qun Li
Location: rem/3a24/3c18
Art Unit: 1648
Wednesday, February 23, 2005

Case Serial Number: 09/664363

From: Deirdre Arnold
Location: Biotech-Chem Library
REM 1A64
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

Search Notes

Please feel free to contact me if you have any questions or would like to amend the search.

Thank you for using STIC services.

Regards,
Deirdre Arnold

This Page Blank (uspto)

From:	Li, Bao-Qun
Sent:	Monday, February 14, 2005 8:40 AM
To:	STIC-Biotech/ChemLib

PLEASE DO THE SEQUENCE HOMOLOGY AND INTERFERENCE OF SEQ ID NO; 22 IN APPLICATION SN.
09,664,363. THANKS.
Bao Qun Li M.D
TC 1600
Art Unit 1648
Tel. 517-272-0904
REM, 3C18
Rm. 3D24

FEB 14 2005

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: February 21, 2005, 09:05:38 ; Search time 15711 Seconds
                                                         (without alignments)
                                                         11565.595 Million cell updates//sec

Title: US-09-664-363-22
Perfect score: 3750
Sequence: 1 TGGGAGGGGGCGTCTTCACAGG...........ACTACCTCACCGTGACCCG 3750

Scoring table: IDENTITY NUC
Searched: Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                1:   gb_ba:*
                2:   gb_htg:*
                3:   gb_in:*
                4:   gb_om:*
                5:   gb_ov:*
                6:   gb_pat:*
                7:   gb_ph:*
                8:   gb_pl:*
                9:   gb_pr:*
                10: gb_ro:*
                11: gb_sts:*
                12: gb_sy:*
                13: gb_un:*
                14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3750	100.0	3750	6	A28157	A28157 PT-NANBH mR
2	3750	100.0	3750	6	AR144051	AR144051 Sequence
3	3729.2	99.4	3750	6	A32203	A32203 NANBHV PT p
4	3319.8	88.5	9359	14	AF113916	AF113916 Hepatitis
5	3312.2	88.3	9410	14	HPCKLS2	D50485 Hepatitis C
6	3302.6	88.1	9410	14	HPCKLR2	D50481 Hepatitis C
7	3297.8	87.9	9573	14	AF208024	AF208024 Hepatitis
8	3297.8	87.9	9573	14	AB049090	AB049090 Hepatitis
9	3294.0	87.9	9379	14	HCVPOLYP	AJ000009 Hepatitis
10	3293	87.8	9374	14	AF207753	AF207753 Hepatitis
11	3292.2	87.8	8642	6	AR488101	AR488101 Sequence
12	3292.2	87.8	8642	6	AX472292	AX472292 Sequence
13	3291.4	87.8	7989	6	AR406043	AR406043 Sequence
14	3291.4	87.8	7989	6	AX036255	AX036255 Sequence
15	3291.4	87.8	7989	12	SSE242652	AJ242652 Hepatitis
16	3291.4	87.8	8001	6	AR406042	AR406042 Sequence
17	3291.4	87.8	8001	6	AX036254	AX036254 Sequence
18	3291.4	87.8	8001	12	SSE242654	AJ242654 Hepatitis
19	3291.4	87.8	8637	6	AR406041	AR406041 Sequence
20	3291.4	87.8	8637	6	AX036253	AX036253 Sequence
21	3291.4	87.8	8637	12	SSE242651	AJ242651 Hepatitis
22	3291.4	87.8	8638	6	AR488121	AR488121 Sequence
23	3291.4	87.8	8638	6	AX472314	AX472314 Sequence
24	3291.4	87.8	8639	6	AR488100	AR488100 Sequence
25	3291.4	87.8	8639	6	AX472291	AX472291 Sequence
26	3291.4	87.8	8649	6	AR406044	AR406044 Sequence
27	3291.4	87.8	8649	6	AX036256	AX036256 Sequence
28	3291.4	87.8	8649	12	SSE242653	AJ242653 Hepatitis
29	3291.4	87.8	9605	6	AX739971	AX739971 Sequence
30	3291.4	87.8	9605	14	HCJ238799	HCJ238799 Hepatitis
31	3291.4	87.8	10690	6	AX739972	AX739972 Sequence
32	3291.4	87.8	11076	6	AR406040	AR406040 Sequence
33	3291.4	87.8	11076	6	AX036252	AX036252 Sequence
34	3289.8	87.7	9386	6	AF165055	AF165055 Hepatitis
35	3289.8	87.7	12305	14	AX937622	AX937622 Sequence
36	3289.8	87.7	12315	6	AX937621	AX937621 Sequence
37	3289	87.7	7065	6	A76577	A76577 Sequence 9
38	3288.2	87.7	8638	6	AR488104	AR488104 Sequence
39	3288.2	87.7	8638	6	AR488105	AR488105 Sequence
40	3288.2	87.7	8638	6	AR488122	AR488122 Sequence
41	3288.2	87.7	8638	6	AX472296	AX472296 Sequence
42	3288.2	87.7	8638	6	AX472295	AX472295 Sequence
43	3288.2	87.7	8638	6	AX472315	AX472315 Sequence
44	3288.2	87.7	8643	6	AR488102	AR488102 Sequence
45	3288.2	87.7	8643	6	AX472294	AX472294 Sequence

ALIGNMENTS

RESULT 1
A28157
LOCUS   A28157   3750 bp   DNA   linear   PAT 04-AUG-1995
DEFINITION   PT-NANBH mRNA fragment from patent GB2239245.
ACCESSION   A28157
VERSION   A28157.1   GI:1248640
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM   synthetic construct
                other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 3750)
  AUTHORS
  TITLE   Post-transfusional non-A non-B hepatitis viral polypeptides
  JOURNAL   Patent: GB 2239245-A 22 26-JUN-1991;
FEATURES   Location/Qualifiers
     source   1..3750
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3630"
                     /note="contig formed by cDNA clones from 3' end"
     CDS   <1..>3750
                     /codon_start=1
                     /transl_table=11
                     /product="PT-NANBH polyprotein"
                     /protein_id="CAA01932.1"
                     /db_xref="GI:1248641"
                     /translation="WEGVFTGLTHVDAHFLSQTKQAGDNFPYLVAYQATVCARAQPP
                     PSWDQMWKCLIRLKFTLRGPTIPLLYRLGAVQNEVTLITHPITKFIMACMSADLEVTST
                     WVLVGGVLAALAAYCLTTGSVVIVGRIILSGRPAIVPDREVLYQEFDEMEECASHLPY
                     IEQGMQLAEQFKQKALGLLQTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYL
                     AGLSTLPGNPAIASLMAFTASVTSPLTTQGTLLNILGGWVAAQLAPPSAASAFVGAG
                     IAGAAVGSIGLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGA
                     LVVGVVCAAILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILS
                     FFSCQRGYKGVWRDGIMQTTCSCGAQITGHVKNGQNSMRIVGPKTCSNMWHGTFPINAY
                     TTGPCTPSPAPNYSRALWRVAAEEYVEVTRVGDPHYVTSMTTNVKCPCQVPAPEFFT
                     EVDGVRLHRYAPACKPLREEVTFQVGLNQYLVGSQLPCEPEPDVAVITSMLTDPSHI
                     TAETAKRRLARGSPPSLASSSASQLSAPSSKATYITTQNDFPDADLIEANLLWRHEMGG
                     DITRVESNKVVILDSFDPLRAEEDEREVSVPAEILRKSKSKFFPAMPAWARPDYNPPL
                     LESWKAPDYVPPVVHGCPLPPTKVPPIPPPRRKRTVVLTESTVSSALAELATKAFGSS
                     EPSAVDSGTATAPPDQPSDDGGAGSDVRSSYSSMPPLEGEPGDPDLSDGSWSTVSEEAG
                     EDVVCCSMSYTWTGALIITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQROKKV"

ORIGIN

```
TFDRLQILDDHYQDVLKEMKAKASTVKAKLLSVEEACKLTPHSAKSKFGYGAKDVRN
LSSKAINHIRSWWEDLLEDTETPIDTTIMAKNEVFCVQPERGRKPARLIVFPDLGVR
VCEKMALYDVVSTLPQAVMGSSYGFQYSPGQRVEFLVNAWKSKKTPMGFAYDTRCPDS
TVTENDIRVEEBSIYQCCDLAPEARQAIRSLTERLYIGGPLTMNSKGQNCGYRRCRASGV
LTTSCGNTLTCYLKASACRAAKLQDCTMLVCCDGLVVICESAGTOEDAASLRVFTEA
MTRYSAPPGDPPQPEYDLELITSCSSNVSVAHDASGKRVYLTRDP"
```

Query Match   100.0%;   Score 3750;   DB 6;   Length 3750;
Best Local Similarity   100.0%;   Pred. No.   0;   Mismatches   0;
Matches 3750;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 TGGGAGGGGGCTCTTCACAGGCCTCACCCACGTGGAATGCCCACTTCCTGTCCCAAACAAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGGAGGGGGCTCTTCACAGGCCTCACCCACGTGGAATGCCCACTTCCTGTCCCAAACAAAG 60

Qy     61 CAGGCAGGAGACAACTTCCCTACCTGGTGGCGTACCAGCTACTTCACCCAACTGTTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGGCAGGAGACAACTTCCCTACCTGGTGGCGTACCAGCTACTTCACCCAACTGTTG 120
```

```
Db   1981  CAAAATGACTTCCCAGACGGCTGACCTTCATCGTGGGCGCATGAGATG

Qy   2041  GGGCGGGGACATTACCCGCGTGGGAGTTCAGAGAACAAAGGTAGTAATCCTGGACTCTTTGAC  2100
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  GGCGGGGGACATTACCCGCGTGGGAGTTCAGAGAACAAAGGTAGTAATCCTGGACTCTTTCGAC  2100

Qy   2101  CCGCTCCGAGCGGAGGAGGATGAGCGGGGAAGTGTCCGTCCCGGGCGAAA  2160
Db   2101  CCGGCTCCGAGCGGAGGAGGATGAGCGGGGAAGTGTCCGTCCCGGGCGAAA  2160

Qy   2161  TCCAAGAAATTCCCACCACGGAGATGGCCCGCCGGCCGCATGGGGACTTCGGCTG  2220
Db   2161  TCCAAGAAATTCCCACCACGGAGATGGCCCGCCGGCCGCATGGGGACTTCGGCTG  2220

Qy   2221  CTGGAGTCCTGGAAGGCCCCGGAACTAGTTCCCTCCAGTGGTGCCCACTGCCA  2280
Db   2221  CTGGAGTCCTGGAAGGCCCCGGAACTACGTTCCCTCCAGTGGTGCCCACTGCCA  2280

Qy   2281  CCTACTAAGACCCCTCCTATACCACCTCCATTTTCCATTTTGTTCTGACAGAA  2340
Db   2281  CCTACTAAGACCCCTCCTATACCACCTCCATTTTCCATTTTGTTCTGACAGAA  2340

Qy   2341  TCCACCGGTGTCTTCTGCCCTGGGGCGGGGCGGGGCTTTGGGTTAGCTTCCGAAACCG  2400
Db   2341  TCCACCGGTGTCTTCTGCCCTGGGGCGGGGCGGGGCTTTGGGTTAGCTTCCGAAACCG  2400

Qy   2401  TCGGCCGGTGGACAGGGGCAGGGCAACCGCCCCTCCTGACCCGGA  2460
Db   2401  TCGGCCGGTGGACAGGGGCAGGGCAACCGCCCCTCCTGACCCGGGA  2460

Qy   2461  GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCTTGAGGGGGGAC  2520
Db   2461  GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCTTGAGGGGGGGAC  2520

Qy   2521  CCCGATCTCAGCGGACGGGGTCTTGGTCTTACCGGTGAGGGACTCGTC  2580
Db   2521  CCCGATCTCAGCGGACGGGGTCTTGGTCTTACCGGTGAGGGACTCGTC  2580

Qy   2581  TGCTGCTCGATGTCCTTACACATGGACAGGGGGCTCTGATCACGCCATTGCCC  2640
Db   2581  TGCTGCTCGATGTCCTTACACATGGACAGGGGGCTCTGATCACGCCATTGCCC  2640

Qy   2641  AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGGTCGCACACAACATTGTCTAC  2700
Db   2641  AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGGTCGCACACAACATTGTCTAC  2700

Qy   2701  GCTACCACATCCCGCAGCCGCAAGCCCAGCAGGCCAGAGAAGGTCACCTTTGACAGACTGCAA  2760
Db   2701  GCTACCACATCCCGCAGCCGCAAGCCCAGCAGGCCAGAGAAGGTCACCTTTGACAGACTGCAA  2760

Qy   2761  ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCAAAGGCAGTTCCACAGTT  2820
Db   2761  ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCAAAGGCAGTTCCACAGTT  2820

Qy   2821  AAGGCTAAGCTTCTATCAGTGAGAAGCTGCAAGCTGCAAGCTGCGAGTCAAGTCGAAGCTGCCAA  2880
Db   2821  AAGGCTAAGCTTCTATCAGTGAGAAGCTGCAAGCTGCAAGCTGCGAGTCAAGTCGAAGCTGCCAA  2880

Qy   2881  TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAAGGCCATTAACCAC  2940
Db   2881  TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAAGGCCATTAACCAC  2940

Qy   2941  ATCCGGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCACCATC  3000
Db   2941  ATCCGGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCACCATC  3000

Qy   3001  ATGGCAAAAAAATGAGGTTTTCTGGCTCCAACCAGAGAGAGGCCGCAAGCCAGCTCGC  3060
Db   3001  ATGGCAAAAAAATGAGGTTTTCTGGCTCCAACCAGAGAGAGGCCGCAAGCCAGCTCGC  3060

Qy   3061  CTTATCGTGTTCCAGACTTGGGGGGTTCCGTGTGTGCCGAGAAAATGGCCCTTCTATGACGTG  3120
Db   3061  CTTATCGTGTTCCAGACTTGGGGGGTTCCGTGTGTGCCGAGAAAATGGCCCTTCTATGACGTG  3120

Qy   3121  GTCTTCACCCTCCCCTCAGGCGTGTGATGGGGCGTCCTCCTGATTTCCAGTATTCTCCTGGA  3180
Db   3121  GTCTTCCACCCTCCCCTCAGGCGTGTGATGGGGCGTCCTCCTGATTTCCAGTATTCTCCTGGA  3180

Qy   3181  CAGCGCGGGTCGAGTTCCTGCTGGTGAAACGGCCCTGGAAATCAAAAGAGAAGACCCCTATGGGGCTTTGCA  3240
Db   3181  CAGCGCGGGTCGAGTTCCTGCTGGTGAAACGGCCCTGGAAATCAAAAGAGAAGACCCCTATGGGGCTTTGCA  3240

Qy   3241  TATGACAACCCCTCGTTTTGACTCACAGACTCACTGAGAATGACATCCGTGTAGAGGAGTCA  3300
Db   3241  TATGACAACCCCTCGTTTTGACTCACAGACTCACTGAGAATGACATCCGTGTAGAGGAGTCA  3300

Qy   3301  ATTTATCAAATGTTGTGACTTGGCCCCCGGAAGCCACAGCCCAGGCCCATAAGGTCGCTCACAGAG  3360
Db   3301  ATTTATCAAATGTTGTGACTTGGCCCCCGGAAGCCACAGCCCAGGCCCATAAGGTCGCTCACAGAG  3360

Qy   3361  CGGCTTTATATCGGGGGGTCCCCTGACTAATTCAAAAAGGGCAGAACTGCGGCTATCGCCGG  3420
Db   3361  CGGCTTTATATCGGGGGGTCCCCTGACTAATTCAAAAAGGGCAGAACTGCGGCTATCGCCGG  3420

Qy   3421  TGCCGCGCGAGACCGGCGCTATGACTAGCGTAGCTGTGCGGTTAATACCCTCACATGTTACTTGAAG  3480
Db   3421  TGCCGCGCGAGACCGGCGCTATGACTAGCGTAGCTGTGCGGTTAATACCCTCACATGTTACTTGAAG  3480

Qy   3481  GCCTCTGCAGCCGGTGTCGAGCTGCAAAGCTCCAGGACTGCACGACTGCGCAGGAGAC  3540
Db   3481  GCCTCTGCAGCCGGTGTCGAGCTGCAAAGCTCCAGGACTGCACGACTGCGCAGGAGAC  3540

Qy   3541  GGCCTTGTCGTTATCGTGAGGAGCGGCGGGGAACCCAGGAGGAGGAGGAGGCCTACGAGTC  3600
Db   3541  GGCCTTGTCGTTATCTCGTGTGGAGGAGCGGCGGGGAACCCAGGAGGAGGAGGAGGCCTACGAGTC  3600

Qy   3601  TTCACCGGAGGCTATGACTAGGTACTCTTGCCCCCCCCCGGGGGGACCCGCCCGCAACCAGAATAC  3660
Db   3601  TTCACCGGAGGCTATGACTAGGTACTCTTGCCCCCCCCCGGGGGGACCCGCCCGCAACCAGAATAC  3660

Qy   3661  GACCTGGAGTTGATAACATCATGCTCCTCCAAATGTGTTCGGTCGGTCGGGTGCGGTGC  3720
Db   3661  GACCTGGAGTTGATAACATCATGCTCCTCCAAATGTGTTCGGTCGGTCGGGTGCGGTGC  3720

Qy   3721  AAAAGGGTATACTACCTTCACCCGTGACCCG  3750
Db   3721  AAAAGGGTATACTACCTTCACCCGTGACCCG  3750
```

RESULT 2
AR144051

LOCUS	AR144051	3750 bp	DNA	linear	PAT 08-AUG-2001

DEFINITION Sequence 22 from patent US 6210675.
ACCESSION   AR144051
VERSION   AR144051.1 GI:15105918
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3750)
  AUTHORS   Highfield,P.Edmund., Rodgers,B.Colin., Tedder,R.Seton. and
            Barbara,J.Anthony.James.
  TITLE   PT-NANB hepatitis polypeptides
  JOURNAL   Patent: US 6210675-A 22 03-APR-2001;
FEATURES   Location/Qualifiers
     source   1..3750
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 3750; DB 6; Length 3750;
Best Local Similarity 100.0%;
Matches 3750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1  TGGGAGGGGCGTCTTCACAGGCGCCTCACCCACCGTGACCCACTTCCTGTCCCCAAAACAAAG  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     1 TGGGAGGGGCGTCTTCACAGACGTCACCCACTTCCTGTCCCAAAACAAAG 60

Qy    61 CAGGCAGGAGGACAACTTCCCCTACCTGGTGGGCGTACCAGGCTACTGTTGTCGCTAGGGC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CAGGCAGGAGGACAACTTCCCCTACCTGGTGGGCGTACCAGGCTACTGTTGTCGCTAGGGC 120

Qy   121 CAGGCCCCACCTCATCATGGGATCAAAATGTGAAGAGTGTCTCATACGGCTAAAGCCTACT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGGCCCCACCTCATCATGGGATCAAAATGTGAAGAGTGTCTCATACGGCTAAAGCCTACT 180

Qy   181 CTCGCGGCGGCGGCCACACCTCCATCATGGGAGTCTCATAACGGCTAAAGCCTACT 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTCGCGGCGGCGGCCACACCTCCATCATGGGAGTCTCATAACGGCTAAAGCCTACT 240

Qy   241 ACCACCCCCATAACCAAAATTCATCATCATGGCATGCATCACCCCTC 300
         ||||||||||||||||||||||||||||||||||||||||
Db   241 ACACACCCCCATAACCAAAATTCATCATCATGGCATGCATCACCCTCACG 300

Qy   301 AGCACCTGGGTGCTGCTGGTGGGCGGGGGGGGGGGGTCCTTGCAGCTTCTGCTTGACAACA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGCACCTGGGTGCTGCTGGTGGGCGGGGGGGGGGGGTCCTTGCAGCTTCTGCTTGACAACA 360

Qy   361 GGCAGGCGTGGTGCTGGTAGGATCATCTTGTCCGGCGGCGGGGGGGGGTATTGCTTGACAACA 420
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGCAGGCGTGGTGCTGGTAGGATCATCTTGTCCGGCGGCGGGGGGGGGTATTGCTTGACAACA 420

Qy   421 AGGGAAGTCCTCTACCAGGAGTTCGATGAGGATGGAAGAGTTCGCGTGCGACCTCCCTTAC 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AGGGAAGTCCTCTACCAGGAGTTCGATGAGGATGGAAGAGTTCGCGTGCGACCTCCCTTAC 480

Qy   481 ATCGAGCAGGGAAGTCAGCAGTTCGAATGACAGTTCAAGCAAAAAAGGGTGCTCCCTTAC 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATCGAGCAGGGAAGTCAGCAGTTCGAATGACAGTTCAAGCAAAAAAGGGTGCTCCCTTAC 540

Qy   541 ACAGCCACACAAAGCAAGCGGAGGCGCCCCTGCTCCCCGTGGTGGAGTCCGGCGAGCCCTT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACAGCCACACAAAGCAAGCGGAGGCGCCCCTGCTCCCCGTGGTGGAGTCCGGCGAGCCCTT 600

Qy   601 GAGACCTTCTGGGCGAAACACATGTGGGAACTTCATCAGCGGGATACATTAGCAGGC 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGACCTTCTGGGCGAAACACATGTGGGAACTTCATCAGCGGGATACATTAGCAGGC 660

Qy   661 TTGTCCACTCTGCCTGGGAATCCCCGATTGCATCATCGCCTTCACAGCCCTCTGTC 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ACTAGCCCCGCTCACCCCAAATCTACCCTCCTGCTTAACATCCTGGGGGGATGGGTAGCC 780

Qy   721 GCCCAACTCGCTCCCCCCAGTGCTGCTTCACCTTTCGTTAGGCGGCGCGCGGGGCGG 840

Db   781 GCCCAACTCGCTCCCCCCAGTGCTGCTTCACCTTTCGTTAGGCGGCGCGCGGGGCGG 840

Qy   841 GCTGTTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTTGGGACATCTTGGCGGCTATCGGAGCA 900
Db   841 GCTGTTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTTGGGACATCTTGGCGGCTATCGGAGCA 900

Qy   901 GGAGTGGCAGGCGGCGCCTCGTGGGCTCCTGGGTGGGCCTTTAAGGTCATGAGCGGCGGAAATGCCC 960
Db   901 GACCTGGTTAACTTACTCCGTCGGCACTCCCATCCCTCTCTCCTGCCCATCCCTCCTG 960

Qy   961 GACCTGGTTAACTTACTCCGTCGGCACTCCCATCCCTCTCTCCTGCCCATCCCTCCTG 1020
Db   961 GACCTGGTTAACTTACTCCGTCGGCACTCCCATCCCTCTCTCCTGCCCATCCCTCCTG 1020

Qy  1021 TGCGCAGCGGATACTCCGTCGGCACCTGGGTCCAGGGAGGGGGCTGTGCAGTGGATGAAC 1080
Db  1021 TGCGCAGCGGATACTCCGTCGGCACCTGGGTCCAGGGAGGGGGCTGTGCAGTGGATGAAC 1080

Qy  1081 CGGCTGATACGGTTCGCCTCCGCGGGGTAACCATGTTTCCCCACGCACTANTGTGCCAGAG 1140
Db  1081 CGGCTGATACGGTTCGCCTCCGCGGGGTAACCATGTTTCCCCACGCACTANTGTGCCAGAG 1140
```

```
Qy  1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTCTCCGACTTACCAACTGTTG 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTCTCCGACTTACCAACTGTTG 1200

Qy  1201 AAGAGGCTCCACCAGTGGATTAACGGAGGACTGCTCCACGCCCCTTGCTCCTCCGGCTA 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AAGAGGCTCCACCAGTGGATTAACGGAGGACTGCTCCACGCCCCTTGCTCCTCCGGCTA 1260

Qy  1261 AGGGATGTTTGGGACTGGAТATGCACAGTTTTTGGGCGCTTCAAGACCTGGCTTCCAGTCC 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 AGGGATGTTTGGGACTGGAТATGCACAGTTTTTGGGCGCTTCAAGACCTGGCTTCCAGTCC 1320

Qy  1321 AAGCTCCTTGCCGCGGATTACCGGGGAGTCCCCTTTTTCTCATGCCAAGCCTACAAGGGG 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 AAGCTCCTTGCCGCGGATTACCGGGGAGTCCCCTTTTTCTCATGCCAAGCCTACAAGGGG 1380

Qy  1381 GTCTCGGCGGGGAGAGGGCATCATGCAGAGCACACCTGCTCAGTGTGGACACAGATCACCGGA 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GTCTCGGCGGGGAGAGGGCATCATGCAGAGCACACCTGCTCAGTGTGGACACAGATCACCGGA 1440

Qy  1441 CATGTCAAAAAACGGTTCCATCAAGCGAТAСАCCACACGGGAGCCCCTGCGACGCGCAAAAC 1500
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CATGTCAAAAAACGGTTCCATCAAGCGAТAСАCCACACGGGAGCCCCTGCGACGCGCAAAAC 1500

Qy  1501 GGAACATTCCCCATCAACGCATACCACCACGGGTGTTGGGCCTAAGACCGTAGTAACAGTGGCAT 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GGAACATTCCCCATCAACGCATACCACCACGGGTGTTGGGCCTAAGACCGTAGTAACAGTGGCAT 1560

Qy  1561 TATTCCAGGGTCTGTGGGCGGGTGCTGAGGAGTACGGCGGGTTCCA 1620
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 TATTCCAGGGTCTGTGGGCGGGTGCTGAGGAGTACGGCGGGTTCCA 1620

Qy  1621 GATTTCCACTACGTGACGAGGCGATGACCACTGACAGTAAAATGCCCGGTGCCAGTTCCA 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GATTTCCACTACGTGACGAGGCGATGACCACTGACAGTAAAATGCCCGGTGCCAGTTCCA 1680

Qy  1681 GCCCCCGGAATTCTTCACAGAAGTGGATGGGGGCTGCACAGGTAGCGTTCGGGGGTGC 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 GCCCCCGGAATTCTTCACAGAAGTGGATGGGGGCTGCACAGGTAGCGTTCGGGGGTGC 1740

Qy  1741 AAACCTTCTCCTTACGGGAGGCCCGGAACCGGATGGTCACATTCCAGGTGGGCTCAACCGACTCACCGAC 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 TCGCAGCTCCCATGCGAGGAGGTCACATTCCAGGTGGGCTCAACCGACTCACCGAC 1800

Qy  1801 TCGCAGCTCCCATGCGAGGAGGTCACATTCCAGGTGGGCTCAACCGACTCACCGAC 1860
Db  1801 CCGCAGCTCCCATGCGAGGAGGTCACATTCCAGGTGGGCTCAACCGACTCACCGAC 1860

Qy  1861 CCCTCCACAATCACAGCCAGCAGACGGCTAAGGCAGTCCCATGGCCAGGACTCCCCTCC 1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 CCCTCCACAATCACAGCCAGCAGACGGCTAAGGCAGTCCCATGGCCAGGACTCCCCTCC 1920

Qy  1921 TTGGGCCAGCTCTTCAGCTAGCCAGTTGTTCTTGCGGCGCCCTTCCTGTGCGGAGGCGCAGCACTTACC 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 TTGGGCCAGCTCTTCAGCTAGCCAGTTGTTCTTGCGGCGCCCTTCCTGTGCGGAGGCGCAGCACTTACC 1980

Qy  1981 CAAAATGACTTCCCAGACGTGACTTCATGAGGCCAACCTTCCGTGTGGCGGCGCATGAGATG 2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 CAAAATGACTTCCCAGACGTGACTTCATGAGGCCAACCTTCCGTGTGGCGGCGCATGAGATG 2040

Qy  2041 GGCGGGGCAGCATTACCCGCGTGGAGTCAGAGAACACAGGTAGTAATCCTCGGACTCTTTCGAC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 GGCGGGGCAGCATTACCCGCGTGGAGTCAGAGAACACAGGTAGTAATCCTCGGACTCTTTCGAC 2100

Qy  2101 CCGCTCCGAGCGGGAGGAGGAGGGATGAGCGGGAAGTGTCCGTCCCGAGATCCTGCGGAAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCGCTCCGAGCGGGAGGAGGAGGGATGAGCGGGAAGTGTCCGTCCCGAGATCCTGCGGAAA 2160

Qy  2161 TCCAAGAAAATTCCCACCACAGCGATGCCCCGCATGGGCACGCGCCCGGATTACAACCCTCCGCTG 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 TCCAAGAAAATTCCCACCACAGCGATGCCCCGCATGGGCACGCGCCCGGATTACAACCCTCCGCTG 2220
```

```
Qy  2221  CTGGAGTCCTGGGAAGGCCCCGGGACTAGCTACGTCCTGTACACTGGCCA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  CTGGAGTCCTGGGAAGGCCCCGGGACTAGCTACGTCCTGTACACTGGCCA  2280

Qy  2281  CCTACTAAGACCCCTCCTATACCACTCACGGAGGAGGAGAAGCAGTTGTTCTGCAGAA    2340
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  CCTACTAAGACCCCTCCTATACCACTCACGGAGGAGGAGAAGCAGTTGTTCTGACAGAA  2340

Qy  2341  TCCACCGTGTCTTCTGCCCTGGGCGGAGCTTGCCACAAGGCTTTCGGTTACCTTCGGAACG   2400
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TCCACCGTGTCTTCTGCCCTGGGCGGAGCTTGCCACAAGGCTTTCGGTTACCTTCGGAACG  2400

Qy  2401  TCGGCCGTGCGACAGGGCGACGGGCCGGAGCTTGCCACAACCTTCGGGTAGCTCTGAAG  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TCGGCCGTGCGACAGGGCGACGGGCCGGAGCTTGCCACAACCTTCGGGTAGCTCTGAAG  2460

Qy  2461  GCAGGATCTGACGTTGAGTCGTAGTCGTAGTCGTATTCCTCCATGCCCCTCCTGACCAACG    2520
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GCAGGATCTGACGTTGAGTCGTAGTCGTAGTCGTATTCCTCCATGCCCCTCCTGACCAACG  2520

Qy  2521  CCCGATCTCAGCGAGCGGGTTCTTGGTCTTACCGTGAGTGAGGAGGGCCCGGGACGTCGTC  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CCCGATCTCAGCGAGCGGGTTCTTGGTCTTACCGTGAGTGAGGAGGGCCCGGGACGTCGTC  2580

Qy  2581  TGCTGCTTCGATGTCCTTACACATGGCAGGGGCTCTGATCATCACGCCATGGTGCGGAGGAA    2640
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TGCTGCTTCGATGTCCTTACACATGGCAGGGGCTCTGATCATCACGCCATGGTGCGGAGGAA  2640

Qy  2641  AGCAAGCTGCCCATCAAACGGGTTGAGCAACTCTTTGCTCGTCGGGTCTCACCACACATGGTCTAC   2700
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  AGCAAGCTGCCCATCAAACGGGTTGAGCAACTCTTTGCTCGTCGGGTCTCACCACACATGGTCTAC  2700

Qy  2701  GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAGAAGGTCACTCTTTGACAGACTGCAA  2760
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAGAAGGTCACTCTTTGACAGACTGCAA  2760

Qy  2761  ATCCTGGACGATCACTCACCAGGAGCGTGCTCAAAGAGAGATGAAAGGCAAGGGCCTTCCACAGTT   2820
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  ATCCTGGACGATCACTCACCAGGAGCGTGCTCAAAGAGAGATGAAAGGCAAGGGCCTTCCACAGTT  2820

Qy  2821  AAGGCTAAGCTTCTATCAGTAGAGGAAGCTTGACGTCGCAAGCTGACGGCCCCACATTCGGCCAAA  2880
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  AAGGCTAAGCTTCTATCAGTAGAGGAAGCTTGACGTCGCAAGCTGACGGCCCCACATTCGGCCAAA  2880

Qy  2881  TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC   2940
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC  2940

Qy  2941  ATCCGGCTCCGTGTGGGAGGACTTGTTGTTGGAAGAGACTGAAACACCAATTGACACCACCATC   3000
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  ATCCGGCTCCGTGTGGGAGGACTTGTTGTTGGAAGAGACTGAAACACCAATTGACACCACCATC  3000

Qy  3001  ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGGGCCGCAAGCCAGCTCGC  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGGGCCGCAAGCCAGCTCGC  3060

Qy  3061  CTTATCGTTGTTCCAGACTTGGGGGGTCCGGTGTGTCGGAGAGGAGGAGTTGGCCCTCTATGAGCGTG   3120
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  CTTATCGTTGTTCCAGACTTGGGGGGTCCGGTGTGTCGGAGAGGAGGAGTTGGCCCTCTATGAGCGTG  3120

Qy  3121  GTCTTCACCCTCCCTCAGCCTGTCGATGGGCTGTGATGGGCTCCTGGATCCAGTATTCTCCTGGA   3180
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  GTCTTCACCCTCCCTCAGCCTGTCGATGGGCTGTGATGGGCTCCTGGATCCAGTATTCTCCTGGA  3180

Qy  3181  CAGCGGGGTCGAGTTCCTGTGAAATCAAAGAGACCCCTATGGGCCTTTGCA  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  CAGCGGGGTCGAGTTCCTGTGAAATCAAAGAGACCCCTATGGGCCTTTGCA  3240

Qy  3241  TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAAATGACACATCCGTGTAGAGAGTCA  3300
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAAATGACACATCCGTGTAGAGAGTCA  3300

Qy  3301  ATTTATCAATGTTGTGACTTGGCCCCCGGAAGCCAGACGAGCCAGAAGTGCGCTCACAGAG  3360
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  ATTTATCAATGTTGTGACTTGGCCCCCGGAAGCCAGACGAGCCAGAAGTGCGCTCACAGAG  3360

Db  3361  CGGCTTTTATATCGGGGGGGTCCCCTGACTAATTCAAAAGGGCAGAACTGGCGGCTATTCCGCCGG  3420

Db  3421  TGCCGCGGCGGAGGGCGGCGGTGCCTGACGACTAGCTGCGGGTAATACCCTCTCACATGTTACTTGAAG  3480

Db  3481  GCCTCTTCGCCAGCCTGTGCGAGCTGCAAAGTCCAGGACTGCACCGGTAATACCCTCACATGTTACTTGAAG  3540

Db  3541  GGCCCTTGTCGTTATCTGTGAGAGCGGGGAACCCAGGAGGACTGCGCAAAAGCTCCAGGAGAC  3600

Db  3601  TTCACGGAGGGCTATGAGTGACTAGGTAACTCTGCCCCCCACCCCCCAATGTGTCGGGTGGC  3660

Db  3661  GACCTGGAGTTGAATAACATCATGCCTCTCCCAATGTGTCGGTGGCGGCGGGGACGATGCATCGGC  3720

Db  3721  AAAAGGTATACTTCACCCGTGACCCG  3750
```

RESULT 3

```
LOCUS       A32203                  3750 bp    DNA     linear   PAT 10-DEC-1996
DEFINITION  NANBHV PT polyprotein fragment.
ACCESSION   A32203
VERSION     A32203.1  GI:1926564
KEYWORDS    .
SOURCE      Non-A, non-B hepatitis virus
  ORGANISM  Non-A, non-B hepatitis virus
            Viruses; unclassified viruses.
REFERENCE   1  (bases 1 to 3750)
  AUTHORS   .
  JOURNAL   Patent: EP 2655990-A 22 21-JUN-1991;
FEATURES             Location/Qualifiers
     source          1..3750
                     /organism="Non-A, non-B hepatitis virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:12440"
     CDS             <1..>3750
                     /codon_start=1
                     /product="partial NANBH-PT polyprotein"
                     /protein_id="CAA02095.1"
                     /db_xref="GI:1926565"
                     /translation="WEGVTTGLTHVDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPP
                     PSWDQMWKCLIRLKLTLLRGPTPLLYRLRIGAVQNEVTLTHPITKFIMACMSADLEVVTST
                     WVLVGGVLAALAAYCLTTGSSVVIVGRIILSGRPAIVPDREVLVQEFDEMEECASHLPY
                     IEQGMQLAEQFKQKALGLLQTATRKQAEARAAPVVESKWRALETFWAKHMWNFISGIQYL
                     AGLSTLPGNPAIASLMAFFTASVTSPLTTQSTLLNILGGWVPPNSLPPSAASAFVGAG
                     IAGAAVGSIGLGKVLVDIIAGYGAVQOMMNRLIAFASRGNHVSPTHYVPESDAAAARVTQILS
                     LVVGGVCAAILRRHVGPGEGAVQWRLIAFASRGNHVSPTHYVPESDAAAARVTQILS
                     DLITTQLLRLHQWINEDGSTPCSGSWLRDVWDWICTVLADFKTWLQSKLLPRLPGVP
                     FFSCQRGRYXGWRWGDGIWQTTCSCGAQITGHVKNGSMRIVGPRKTCSNMWHGTPINAY
                     TTGPCTPSPAPNYSRALWRVAAEEYVEVTRVGDFHYVTGMTDNVKCPCQVPAPEFFT
                     EVDGVRLHRYAPACKPLLREEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHI
                     TAETAKRRLARGSPPSLASSSAAQLSASSSKATYIITONDFPDDADLIEANLLWRHEMGG
                     DITRVESENKVVILDSFDPLRAAEEREVSVPAEILRKSKKFPPAMPAWARPDYNPPL
                     LESWKAPDYVPPVVHGCPLPPTKTPPIPPRRKRTVVLTESTVSSALALEATKAFGSS
                     EPSAVDSGTATAPPDQPSDDGGAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSSEAG
                     EDVVCCSMSYTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKV
                     TFDRLQILDDHYQDVLKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRN
                     LSSKAINHIRSVWEDLLEDTETPIDTTIMAKNEVFCVQPERGGRKPARLIVFPDLGVR
```

ORIGIN

Query Match   99.4%;   Score 3729.2;   DB 6;   Length 3750;
Best Local Similarity   99.7%;   Pred. No. 0;   Mismatches 13;   Indels 0;   Gaps 0;
Matches 3737;   Conservative   0;   Mismatches   13;   Indels   0;   Gaps   0;

```
VCEKMALYDVVSTLPQAVMGSSYGFQYSPGQRVEFLVNAWKSKKTPMGFAYDTRCFDS
TVTENDIRVEEBSIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGV
LTTSCGNTLTCYLKASAACRAAKLQDCTMLVCGDGLVVICESAGTQBDAASLRVFTEA
MTRYSAPPGDPPEYDLELIISCSSSNVSVAHDASGKRVYLTRDP"
```

```
Qy      1  TGGGAGGGGGTCTTCTTCACAGCCCTCACCCCACGTGGATGTGCCCACTTCCTGTCCCAAACAAAG    60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  TGGGAGGGGGTCTTCTTCACAGCCCTCACCCCACGTGGATGTGCCCACTTCCTGTCCCAAACAAAG    60

Qy     61  CAGGCAGGAGGAGACAACTTCCCCTACCTGGTGTGGCGTACCAGGCTACTGTGTTGTGCGGCTAGGGCC   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CAGGCAGGAGGAGACAACTTCCCCTACCTGGTGTGGCGTACCAGGCTACTGTGTTGTGCGGCTAGGGCC   120

Qy    121  CAGGCCCCACCTTCCATCATGGGATCAAAATGTGGAAGTGTTCATACGGGCTAAAGCCTACT   180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CAGGCCCCACCTTCCATCATGGGATCAAAATGTGGAAGTGTTCATACGGGCTAAAGCTTACT   180

Qy    181  CTGGGCGGCGGCCAAACACCCTTGCTGTATAGGCTGGGGAGCCCTTCCAAAAACGGAGGTCACCCCTC   240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  CTGCGGCGGGGCCAAACACCCTTGCTGTATAGGCTGGGGAGCCCTCCAAAAACGGAGGTCACCCTC   240

Qy    241  ACACACCCCATAACCAAAATTCATCATGCATGTCAGCCACCTGGGAGGTCGTCACG   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  ACACACCCCATAACCAAAATTCATCATGGCATGTCAGCCACCTGGGAGGTCGTCACG   300

Qy    301  AGCACCTGGGGTGCTGGGTGGGTGGGGGGCGGGGGGGGGGGGGGGGGGGTGTCACCGGA   360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  AGCACTCGGGGTGCTGGGTGGGTGGGGGGCGGGGGGGGGGGGGGGGGGGTGTCACGGGA   360

Qy    361  GGCAGCGTGGTCATTGTGGGGTAGGAATCATCTTTGTTCCGGGTAGGGATCATCTTTGCTGCTTAGTGCTTGACACAACA   420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GGCAGCGTGGTCATTGTGGGGTAGGAATCATCTTTGTTCCGGGTAGGGATCATCTTTGCTGCTTAGTGCTTGACAACA   420

Qy    421  AGGGAAGTCCTCTACCAGAGAGTTCGATGAGATGAGGAAGAGTGCGGTGCGTGCGCACCTCCCCTTAC   480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AGGGAAGTCCTCTACCAGAGAGTTCGATGAGATGAGGAAGAGTGCGGTGCGTGCGCACCTCCCTTAC   480

Qy    481  ATCGAGCAGGGAAATGCAGCTCGCGGCCCAGCACGTTCAAGCAAAAGGCCTCGGGTTGCTGCAG   540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  ATCGAGCAGGGAAATGCAGCTCGCGGCCCAGCACGTTCAAGCAAAAGGCCTCGGGTTGCTGCAG   540

Qy    541  ACAGCCACCAAGCAAGCGGAGGCCCGCTGCTCCCCGTGGGTGAGTCAAGTGGGCGAGCCCCTT   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  ACAGCCACCAAGCAAGCGGAGGCCCGCTGCTCCCCGTGGGTGAGTCAAGTGGGCGAGCCCCTT   600

Qy    601  GAGAGCCTTCGGGAAACACAGTGGGAACTTCATCACGGCGGGAATACAGATGTACTTTAGCAGGC   660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GAGAGCCTTCGGGAAACACAGTGGGAACTTCATCACGGCGGGAATACAGATGTACTTTAGCAGGC   660

Qy    661  TTGTCCACTCTGCTGGGAATCCCGCGGATTGCATCACTGATGAGGGCTTCACAGCCTCTGTC   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  TTGTCCACTCTGCTGGGAATCCCGCGGATTGCATCACTGATGAGGGCTTCACAGCCTCTGTC   720

Qy    721  ACTACCCCGTCACCCAATCTACCCTCTCCTGCTTAAACATCCTGGGGGATGGGTGCCG   780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  ACTACCCGTCACCCAATCTACCCTCTCCTGCTTAAACATCCTGGGGGATGGGTGCCG   780

Qy    781  GCCCAAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTAGGCGCCGGCCATTGCTGGTGCGG   840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  CCCAAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTAGGCGCCGGCCATTGCTGGTGCGG   840

Qy    841  GCTGTTGGGCAGCATAGGGCCTTGGGGAAGGTGCTTGTGTGGACACATCTTGGCGGCTATGGAGCA   900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  GCTGTTGGGCAGCATAGGCCTTGGGGAAGGTGCTTGTGTGGACACATCTTGGCGGCTATGGAGCA   900

Qy    901  GGAGTGGGCAGGCGGCGCTCGTGGCCCTTTAAGGTCATGAGAGCGGCGGCGAAATGCCCTCCACGGAG   960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GGAGTGGGCAGGCGGCGCTCGTGGCCCTTTAAGGTCATGAGAGCGGCGGCGAAATGCCCTCCACGGAG   960
```

```
Qy    961  GACCTGGTTAACTTCACTTCCTGCCATCCTCTCTTGGTGTCCGTCGTGGGGGTCGTG   1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  GACCTGGTTAACTTACTTCCTGCCATCCTCTCTTGGTGTCCGTCGTGGGGGTCGTG   1020

Qy   1021  TGGCGACGGCGGATTACTCGGCGTCCGGCACGTGGGGTCGTGTCCGGCGGCACGTGGGGTCAGTGAGTGCGTCAGTGGGATGAAC   1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  TGGCGACGGCGGATTACTCGGCGTCCGGCACGTGGGGTCGTGTCCGGCGGCACGTGGGGTCAGTGAGTGCGTCAGTGGGATGAAC   1080

Qy   1081  CGGCTGGATAGCGGTTCGCCCTCGGCCGGGGTAACCATTGTTTCCCCACCGGCACTATGTGCCAGAG   1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  CGGCTGGATAGCGGTTCGCCCTCGGCCGGGGTAACCATTGTTTCCCCACCGGCACTATGTGCCAGAG   1140

Qy   1141  AGCGGACGGCCCGCAGCACGTGTCACTCAGATCCTTCTCCGACCTTAGTCACCCAACTGTTG   1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  AGCGGACGGCCCGCAGCACGTGTCACTCAGATCCTTCTCCGACCTTAGTCACCCAACTGTTG   1200

Qy   1201  AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCAGCGCCCCTGCTCCCGGCGGCTA   1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCAGCGCCCCTGCTCCCGGCGGCTA   1260

Qy   1261  AGGGATGTTCGGGACTGGATTGCACCAGTTGTGATGCTCAAGACCTGGCTCCAGTCC   1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  AGGGATGTTCGGGACTGGATTGCACAGTTGTGATGCTCAAGACCTGGCTCCAGTCC   1320

Qy   1321  AAGCTCCTGCGCGGATTACGGGGAGTCCCCTTTTTCTCATGCCAACGTGGGGTACAAAGGGG   1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  AAGCTCCTGCGCGGATTACGGGGAGTCCCCTTTTTCTCATGCCAACGTGGGGTACAAAGGGG   1380

Qy   1381  GTCTCGGGCGGGGGAGGACGGGCATCATCGCAGACCACCTGCTCATGTGTGGACCGGCACCGGA   1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  GTCTCGGGCGGGGAGGACCGGCATCATCGCAGACCACCTGCTCATGTGTGGACCGGCACCGGA   1440

Qy   1441  CATGTCAAAAACGGTTCCATGAGGATCGTTGGGGCTTAAGACCTGTAGTAACATGGCAT   1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  CATGTCAAAAACGGTTCCATGAGGATCGTTGGGGCTTAAGACCTGTAGTAACATGGCAT   1500

Qy   1501  GGAACATTCCCCATCAAACGCATACACCACCGCGGGGCCCCTGCACGCGCCAAAC   1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  GGAACATTCCCCATCAAACGCATACACCACCGCGGGGCCCCTGCACGCGCCAAAC   1560

Qy   1561  TATTCCGGGGCTGTGGGGCTGTGGCTGTTGGAGGAGTTACGGGGGGTGGGGG   1620
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  GTCTCGGCGGGAGACCGTCGGTGGCGGTTACGGTGGAGTTACGGGGTCGGGG   1620

Qy   1621  GATTTCCACTACGTGACGAGCGATGACCACTGACAAGGTAAAAATGCCCGGGCCAGTTCCA   1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  GATTTCCACTACGTGACGAGCGATGACCACTGACAAGGTAAAAATGCCCGGGCCAGTTCCA   1680

Qy   1681  GCCCCCGGAATTCTTCACAGAAGTGGATGGGGTGCGGCGCTGACAGGTACGGCTGC   1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  GCCCCCGGAATTCTTCACAGAAGTGGATGGGGTGCGGCGCTGACAGGTACGGCTGC   1740

Qy   1741  AAACCTTCCTCTACGGGGAGGAGGTCACAATTCCAGGTGCAAACTACCTGGTTGGG   1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  AAACCTTCCTCTACGGGGAGGAGGTCACAATTCCAGGTGCAAACTACCTGGTTGGG   1800

Qy   1801  TCGCAGCTCCCATGCGAGCCCGGAGTGTAGCAGTTGTCTCCATGCCTCACCGGAC   1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  TCGCAGCTCCCATGCGAGCCCGGAGTGTAGCAGTTGTCTCCATGCCTCACCGGAC   1860

Qy   1861  CCCTCCCACATCACAGCAGAGACGGCTAAGCGCTGGCAGCGGGGTCTCCCCCTCC   1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  CCCTCCCACATCACAGCAGAGACGGCTAAGCGCTGGCAGCGGGGTCTCCCCCTCC   1920

Qy   1921  TTGGCCAGCTCTTCAGCTAGCCAGTTGTTCTCGCCGTTCCTCCGAAGGCGACATACAATTACC   1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  TTGGCCAGCTCTTCAGCTAGCCAGTTGTTCTCGCCGTTCCTCCGAAGGCGACATACAATTACC   1980

Qy   1981  CAAAATGACTTCCCAGAACGTTGAACTTCATCGAGGCCAAACCTTCCTCGTGTGGGGGGGCCATGAGATG   2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  CAAAATGACTTCCCAGAACGTTGAACTTCATCGAGGCCAAACCTTCCTCGTGTGGGGGGGCCATGAGATG   2040
```

```
Qy  2041  GGCGGGGACATTACCCGGCGTGGGAGTGTCAGAGTACTAATCCTGGACTCTTTCGAC  2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  GGCGGGGACATTACCCGGCGTGGGAGTGTCAGAGTACTAATCCTGGACTCTTTCGAC  2100

Qy  2101  CCGCTCCCGAGCGGAGGAGGGAGGGAGGAGTGTCCGTCCGGGCGGGAGATCCTGGGGAAA  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  CCGCTCCCGAGCGGAGGAGGGAGGGAGGAGTGTCCGTCCGGGCGGGAGATCCTGGGGAAA  2160

Qy  2161  TCCAAGAAATTCCCACCAGGATGCCCGCCCGATTACAACCCTCCGCTG  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  TCCAAGAAATTCCCACCAGGATGCCCGCCCGATTACAACCCTCCGCTG  2220

Qy  2221  CTGGAGTCCTGGAAGGGCCGGGACTACGGTCCCTCCAGTGGTCACCCACTGCCA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  CTGGAGTCCTGGAAGGGCCGGGACTACGGTCCCTCCAGTGGTCACCCACTGCCA  2280

Qy  2281  CCTACTAAGACCCCCTCCTATACCACCTCCACGGGGGGGGGAAGCACGTTGTTCTGACAGAA  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  CCTACTAAGACCCCCTCCTATACCACCTCCACGGGGGGGGGAAGCACGTTGTTCTGACAGAA  2340

Qy  2341  TCCACCGTGTCTTCTGCCCTGGCGGGGACGCTTGCCACAAAAGGCTTTGGGTAGCTCCGAACCG  2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TCCACCGTGTCTTCTGCCCTGGCGGGGACGCTTGCCACAAAAGGCTTTGGGTAGCTCCGAACCG  2400

Qy  2401  TCGGCCGTCGACACGGGCACGGCAACCGCCTCCTGACCAACCCTCGACGACGGGGGA  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TCGGCCGTCGACACGGGCACGGCAACCGCCTCCTGACCAACCCTCGACGACGGGGGA  2460

Qy  2461  GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCTTTGAGGGGGGAGCCCGGGGGAC  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCTTTGAGGGGGGAGCCCGGGGGAC  2520

Qy  2521  CCCGATCTCAGCGACGGGGTCTTGGGTCTACCGGTGAGGTGAGGAGGACGTGTC  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CCCGATCTCAGCGACGGGGTCTTGGGTCTACCGGTGAGGTGAGGAGGACGTGTC  2580

Qy  2581  TGCTGCTTGAGTGTTCTACACATGACACGCGCCTCTGATCACGCCATGGCGTGCGAGGAA  2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TGCTGCTTGAGTGTTCTACACATGACACGCGCCTCTGATCACGCCATGGCGTGCGAGGAA  2640

Qy  2641  AGCAAGCTGCCCATCAACGCGGTTGAGCAACTCTTTGCTCGGTCACCAACATGGTCTAC  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  AGCAAGCTGCCCATCAACGCGGTTGAGCAACTCTTTGCTCGGTCACCAACATGGTCTAC  2700

Qy  2701  GCTACCACATCCCGCAGCGCGCAGCGGCGCAGAAGAAGGTCACTTTTTGACAGACTGCAA  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  GCTACCACATCCCGCAGCGCGCAGCGGCGCAGAAGAAGGTCACTTTTTGACAGACTGCAA  2760

Qy  2761  ATCCTGGACGATCACTACCAGGACAGACGTGTCTCAAGAGAATGAAGGGGAAGGCGGTCCACAGTT  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  ATCCTGGACGATCACTACCAGGACAGACGTGTCTCAAGAGAATGAAGGGGAAGGCGGTCCACAGTT  2820

Qy  2821  AAGGCTAAGCTTCTTATCAGTAGGAGGAAGCCTGCAAGCTGACGCGCCACACATTCGGCCAAA  2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  AAGGCTAAGCTTCTTATCAGTAGGAGGAAGCCTGCAAGCTGACGCGCCACACATTCGGCCAAA  2880

Qy  2881  TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCGCAAAGGCCATTAAACCAC  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCGCAAAGGCCATTAAACCAC  2940

Qy  2941  ATCCGCTCCGTGTGGGAGACTTGTTTGGAAGACACTGAAACACCACCATC  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  ATCCGCTCCGTGTGGGAGACTTGTTTGGAAGACACTGAAACACCACCATC  3000

Qy  3001  ATGGCAAAAAATGAGGGTTTTCTCGGTCCAACCACAGAGAGCCCGCAAGCCAGCTCGC  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  ATGGCAAAAAATGAGGGTTTTCTCGGTCCAACCACAGAGAGCCCGCAAGCCAGCTCGC  3060

Qy  3061  CTTATCGTTGTTTCCCGAGACTTGGGGGGGTCCGTTGTGTGGGAGAGAAAAATGGCCCTTCTATGACGTG  3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  CTTATCGTTGTTTCCCGAGACTTGGGGGGGTCCGTTGTGTGGGAGAGAAAAATGGCCCTTCTATGACGTG  3120

Qy  3121  GTCTCCACCCTCCCTCAGCGTGTGATGGGCGGATTCCAGCGTATTCTCCTGGA  3180
```

```
Db  3121  GTCTCCACCCTCCCTCAGGCGTGTGATGGGGGCTCCTCGTGATGGGGGCTTCCTCGTACGGGATTCCAGTGATTCCTCCTGGA  3180

Qy  3181  CAGCGGGGTCGAGTTCCTGGGGTCGAAACCGGGTGAAACGGCCCTTGGGAAATCAAAGAGAGACCCCTATGGGGCTTTGCA  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  CAGCGGGGTCGAGTTCCTGGGGTCGAAACCGGGTGAAACGGCCCTTGGGAAATCAAAGAGAGACCCCTATGGGGCTTTGCA  3240

Qy  3241  TATGACACCCGCGTTTTGACTCAACAGTCACTGAGGAATGACATCCGTGTAGAGGAGGTCA  3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TATGACACCCGCGTTTTGACTCAACAGTCACTGAGGAATGACATCCGTGTAGAGGAGGTCA  3300

Qy  3301  ATTTATCAAATGTTGTGACTTGGCCCCCGAAGCCAGAGGCCATAAGGTCGCGCTCACAGAG  3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  ATTTATCAAATGTTGTGACTTGGCCCCCGAAGCCAGAGGCCATAAGGTCGCGCTCACAGAG  3360

Qy  3361  CGGCTTTATATCGGGGGGTCCCCTGACTAATTCAAAAGGCAGAACTGCGGTATGGCCGG  3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  CGGCTTTATATCGGGGGGTCCCCTGACTAATTCAAAAGGCAGAACTGCGGTATGGCCGG  3420

Qy  3421  TGCCGGCGGACGGGCGTGCTGAACGGACTAGCTGCGGTAATACCCTACCACATGTTACTTGAAG  3480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  TGCCGGCGGACGGGCGTGCTGAACGGACTAGCTGCGGTAATACCCTACCACATGTTACTTGAAG  3480

Qy  3481  GCCTCTGCAGCCTGTCGGAGCTGCAAAGCTCAGGACTGCACGGATGCTTCGTGTCGGTGCGGGGAGAC  3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  GCCTCTGCAGCCTGTCGGAGCTGCAAAGCTCAGGACTGCACGGATGCTTCGTGTCGGTGCGGGGAGAC  3540

Qy  3541  GGCCTTGTCGTTATCTGTTGAGAGCGCGGGAACCCAGGAGGAGGACGCCTAGAGAGTC  3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  GGCCTTGTCGTTATCTGTTGAGAGCGCGGGAACCCAGGAGGAGGACGCCTAGAGAGTC  3600

Qy  3601  TTCACGGGAGGGCTATGACTAGGTATCTGCCCCCCCGGGGGACCCGCCCAACCAGAGAATAC  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  TTCACGGGAGGGCTATGACTAGGTATCTGCCCCCCCGGGGGACCCGCCCAACCAGAGAATAC  3660

Qy  3661  GACCTGGAGTTGATAAACATCATGCTCCTCCAAATGTGTCGCGGGGCACGGATGCATCTGGC  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  GACCTGGAGTTGATAAACATCATGCTCCTCCAAATGTGTCGCGGGGCACGGATGCATCTGGC  3720

Qy  3721  AAAAGGGTATACTACCTCACCCGTGACCCG  3750
          |||||||||||||||||||||||||||||||
Db  3721  AAAAGGGTATACTACCTCACCCGTGACCCG  3750
```

```
RESULT 4
AF313916
LOCUS       AF313916                9359 bp    mRNA    linear   VRL 01-JAN-2002
DEFINITION  Hepatitis C virus polyprotein mRNA, partial cds.
ACCESSION   AF313916
VERSION     AF313916.1  GI:18027684
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9359)
  AUTHORS   Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
  TITLE     Characteristics of Hepatitis C viral genome associated with disease
            progression in a homogeneous patient population
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9359)
  AUTHORS   Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2000) Medicine, National University of Ireland,
            Cork, Hepatitis C Unit, Clinical Sciences Building, Cork University
            Hospital, Cork, Ireland
FEATURES             Location/Qualifiers
     source          1..9359
                     /organism="Hepatitis C virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11103"
                     /note="isolated from source plasma identified as the
                     single source of the iatrogenic infection of Rhesus
```

negative women in Ireland from May 1977-November 1978;
exposure was through HCV 1b contaminated anti-D
immunoglobulin"

```
5'UTR    1..329
CDS      330..>9359
         /codon_start=1
         /product="polyprotein"
         /protein_id="AAL55821.1"
         /db_xref="GI:18027685"
         /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRGPRL
         GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
         SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
         GVNYATGNLPGCSFSIFLLALLSCLTIPASAIEVRNVSGVYHVTNDCSNASIVYEAAD
         MIMHTPGCVPCVRENNSSRCWVALTPTLAARNSSIPTTTIRRHVDLLVGTAAFCSAMY
         VGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRWAWDMMMNWSPTTALVV
         SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMAGNWAKVLIVMLLFAGVDGETTHVTGGSA
         AHTTRGLATLFSPGPSQNIQLINTNGSWHINRTALNCNDSLNTGFLAALFYAHRFNAS
         GCPERMASCRPIDKFAQGWGPIYAKEPLSLDQKPYCWHYAPQPCGIVPASQVCGPVYC
         FTPSPVVVGTTDRSGVPTYSWGENETDVLLLNNTRPPRGNWFGCTWMNGTGFTKTCGG
         PPCDIGVGNDLIICPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
         TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQVLPCSFTTL
         PALSTGLILHHRNIVDVQYLYGIGSAAVSYAIKWEYVLLLFLLADARVCACLWMMLLI
         IQAEAALENLVVLNAASVAGTHGILSFLVFFCAAWYIKGRLVPGAAYAFYGWPLLL
         LILALPPRAYAMDREMAASCGGAVFIGLALLTLSPHYKVFLARLIWWLQYFITRAEAH
         LQVWIPLNVRGRDAIILLMCAIHPELEIPAITKILLAILIGPMLIVLQYPYFVR
         AHGLIRACVLAREVAGGHYVQVALMRLAALTGTYVYDHLTPLRDWAHEGLRDLAVAVE
         PVVFSDMETKIITWGADTPAACGDIIIGLPVSARRGREILLGPADSLEGQWRLLAPIT
         AYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGTK
         TLAGGKPITQMYTNVDQDLVGWLAPGGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
         DSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMETT
         MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
         GFGAYMSKAHGVDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
         TDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNIGEIPFYGK
         AIPIETIKGGRHLIFCHSKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVVVVA
         TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTTVPQDAVSRSQRRGRTGRG
         RTGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQ
         DHLEFWEGVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
         LIRLKPTLHGPTPLLYRLGAVQNEVILTHPITKFIMACMSADLEVATSTWVLVGGVLA
         ALAAYCLTTGSVIVGRIILSGKPAIVPDREVLYQQFDEMEECATHLPYIEQGMQLAE
         QFKQKALGLLQTATKQAEAAAPVVESKWRALESFWAKHMWNFISGIQYLAGLSTLPGN
         PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSI
         GLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAA
         ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
         RLHQWINEDCSTPCSGSWLRDWDWICIVTLTDDKTWLQSKLLPRLPGVPFLSCQRGYK
         GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSP
         APNYSRALWRVTAEEYVEVTQVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHR
         YAPPCKPLLREEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAERAKRRL
         ARGSPPSLASSSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGGNITRVESEN
         KVILDSFEPLRAEEDEREVSVPAEILRRSRKFPRAMPIWARPDYNPPLVEEWKDPDY
         VPPVHGCPLPPKAPPIPPPRRKRTVVLTESTVSSALAELATKTFGSSESSAVDSGT
         ATAPPDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEASEDVVCCSMS
         YTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGQRKKVTFDRLQVLD
         DHYRDVLKEMKAKASTVKAKLLPVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHI
         LSWKDLLEDTEPTDTITIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
         VVSTLPQAVMGSSYGFQYSPGQRVEFLVNAWKKKKNPMGFAYDTRCFEDSTVESDIRV
         EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
         TCYLKASAACRAAKLQDCTMLVCGDDLVVICESAGTXEDAASLRVFTEAMTRYSAPPG
         DPPKPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETAHTPVNSWL
         GNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
         GLSAFSLHSYSPGEINRVAACLRKLGVPPLRVWRHRARSVRAKLLSQGGRAANCGKYL
         FNWAVRTKLKLTPIPAASQLDLSSWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGV
         GIYLLPNR"
```

ORIGIN

```
Query Match   88.5%; Score 3319.8; DB 14; Length 9359;
Best Local Similarity   92.8%; Pred. No. 0; Mismatches 269; Indels 0; Gaps 0;
Matches 3480; Conservative 0; Mismatches 269; Indels 0; Gaps 0;
```

Qy   1 TGGGAGGGGCGCTCTTCACAGGGCCTCACCCAGTGGTGACCCCAACGAAAG   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5001 TGGGAGGGGCGCTCTTCACAGGGCCTCACCCACAATAGATGCCCACACG   5060

Qy   61 CAGGCAGGAGGACAACTTCCCTACTGGTGGTACCAGGCTACTGTGTCGCCTACTAGGGCT 120
           ||||||||||||||||||||||||||| ||||| |||||||||||||||| || |||||
Db   5061 CAGGCAGGAGGACAACTTCCCTACTTGGTTAGCATACCAGGCTACCGGTGTCGCCTACT 5120

Qy   121 CAGGCCCCACCTTCCATCATGGGATCAAAATGTGGAAGTGTCTAAAGCCTACT 180
           ||| ||| |||||||||||| ||||||||||||||||||||||||| |||||
Db   5121 CAGGGCTTCCACCCCATCCCGTGGGACACAAAATGTGGAAGTGTCTCATGGCAG 5180

Qy   181 CTGCGCGGGGCCAACACCCTTGCTTGTTATAGGCTGGGAGCCCTCCAACCTC 240
           |||| ||||||||||||||||||||||||||||||||||||||||||| |||
Db   5181 CTGCAGCAGGGCGGCCAACACCCTGCTTGTATAGGCTGGGCTAGGAGGTCATCCTC 5240

Qy   241 ACACACCCCATAACCAAAATTCATCATGGCATGCAGCCGTCCAAAATGAGG 300
           ||||||| ||||||||||||||||| |||||||||||||||||| |||||
Db   5241 ACACACCCCATTAACCTAAATTCATCATGGGCATGCAGCCGTCGACCTGGAGGTCGTCACG 5300

Qy   301 AGCACCTGGGTGCTCAGGGTGGGGTCCTTGCAGCTTCCTTTAC 360
           ||||||||||||||||||||||| |||||||||||||||||||
Db   5301 AGCACCTGGGTGCTAGTGAGGCGGGAGTCCTTGCCCTCTGGGTAATTGGCCTGACAACA 5360

Qy   361 GGCACGCTGGTCATTGTGGGTAGGAGCTCGGGGCGGGGCCGCGGGTTCCCGAC 420
           ||| |||||||||||||||||||||||| ||||| |||| |||||||| || |
Db   5361 GGCACGCTGGTCATTGTGGGTAGGCGGAGGTCGTGGTCAGGGCTCGGGGTTGTCCGAC 5420

Qy   421 AGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCGCGCACCTCCTTAC 480
           ||||| ||||||||| ||| |||||||| || |||||||||| ||||| ||||||
Db   5421 AGGGAAGTCCTCTACCAGCAGTTCGATGAGATGGAAGAGTGCGCCACACTCCCTTAC 5480

Qy   481 ATCGAGCAGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAGCGCTGGGTCTG 540
           |||||||| ||||||| ||| ||||||||| | ||||| ||| |||||||||||
Db   5481 ATCGAACAGGGAATGCAGCTCGCCGAACAATTCAAGCAGAAGGCGCTGGGTCTG 5540

Qy   541 ACAGCCACCAAGCAAGCAGAGGCGGCTGCCGCTGCTCCCGGTGGTGAGGCCCTT 600
           |||||||||||||||||||||| ||||||| |||||||||||||||||||||||
Db   5541 ACAGCCACCAAGCAAGCAGAGGCTCAAAGTGCCGCTGCTCCCGGTGGTTGAGGCCCTT 5600

Qy   601 GAGAGCTTCTGGGGCGGAATCCCCGGATTGCAATCACTCAGCCTTAGCAGG 660
           || ||||||||||| |||||| |||||||| ||||| | || ||||||||
Db   5601 GAGAGCTTCTGGGCGAAAACATATGCGGAATTTCAACCTACTACTTACGCAGG 5660

Qy   661 TTGTCCACTCTGCCCTGGGAATCCCCGGATTGCCCCGGAATTCAGCCCTGTC 720
           |||| ||||||||||||||||||||| |||||| |||| ||| ||||| |||
Db   5661 TTGTCCAACTCTGCCTGGGAATCCCCCGGAATCACTGAATGGCCACTTCATTC 5720

Qy   721 ACTAGCCCGCTCACCCAAATCTACCCTCCTGCTTAACATCCTCGGGGGATGGGG 780
           || ||||||||||||||||||||||| ||| ||||||||| || ||||||||
Db   5721 ACCAGCGCCGCTTACCACCAACATACCCTCCTTACAGCATTGGGGGATGGGGCT 5780

Qy   781 GCCCCAACTCGGCTCCCCCCAGTCGCTGCTTCAGCTTTCGGTAGGCGGCTGTG 840
           ||||||||| || || ||| ||| || ||| |||| ||||||| | ||||||
Db   5781 GCCCAACTCGCCCCCCACGGGTCGCTTGTCGCTTTCGTAGGCCCGTCGGCGG 5840

Qy   841 GCTGTTGGCAGGATACTGCGGCGGTGGGGTCCAGGGAGGTGCTTGTGGAAGCA 900
           ||||||| || |||||||||| || |||||||| |||| ||| ||| |||||
Db   5841 GCCGGTTGGCAGGTTCGCAGGCGCGTCGGGCCCCAGGGCGGTTTATGGAGCG 5900

Qy   901 GGAGGTGGCAGGCGCGCCCTTGGCCCTTAAGGTCATCAGCGCGGCCAATGTCCC 960
           |||| | || ||||||||||| ||||||| |||||||||||||||||| |||
Db   5901 GGGGGTGGCAGGCGCGCCCTTTGGCCTTTAAGGTCATGAGTGGCGAGATGGCTTC 5960

Qy   961 GACCTCGGTTAACTTACTCCCCGTGCCCATCCTCTCTGGTGCGCTCTGGTCGGTTCG 1020
           |||||||||| | ||||| |||||| ||||||||| ||| ||||| | || |||
Db   5961 GACTTCGGTCAACTTCTGTCTCCCCATCCTCTCCTTGGCTCGCGTGCCGGGGGTTAGTG 6020

Qy   1021 TGGCGAGCGGATACTGGGTGGGGTCCAGGGGACGTCCAGGGGATGAAC 1080
           ||||| |||||| ||| || || |||| || | |||||| || ||||
Db   6021 TGGCGAGCAATACTGCGGTCGGCATGTGGGCCCAGGGGCTTGCAGTGGGATGAAC 6080

Qy   1081 CGGCTGATAGGGTTCGGTTCCGCTGCTCGGCGGTCAACCATGTTTCCCCCACCGGAG 1140
           ||| |||| || || ||| |||| || ||| ||||| |||||||| | || |||||
Db   6081 CGGCTGATAGGCGTTCGCTTGCCCTGCGGGGTCTCCCCCGACCATTAGTTGCCCGAG 6140

Qy   1141 AGCGAGCGCCGCAGCAGCGTGCACTCAGATCCTCTCCGGACTTAGTATCACCCGACTGTG 1200
           ||||| | ||||||||||||||| |||| ||| | |||||| |||||||||||| |||
Db   61 CAGGCAGGAGGACAACTTCCCTACTGGTGGTACCAGGCTACTGTGTCGCCTACTAGGGCT 120

(Page contains a two-column DNA sequence alignment between query "Qy" and database "Db" sequences, with match bars. The nucleotide sequence detail is not legibly reproducible.)

RESULT 5
HPCK1S2

LOCUS   HPCK1S2   9410 bp   RNA   linear   VRL 10-FEB-1999
DEFINITION   Hepatitis C virus (strain HCV-1b, clone HCV-K1-S2), complete genome
             sequence.
ACCESSION   D50485
VERSION   D50485.1 GI:1030704
KEYWORDS   polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
             NS5B; envelope protein; non-structural protein; IFN-sensitive;
             interferon-sensitive; interferon-resistant; IFN-sensitive;
             IFN-resistant; ISDR; interferon sensitivity determining region;
             HVR; hypervariable region.
SOURCE   Hepatitis C virus
  ORGANISM   Hepatitis C virus
             Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
             Hepacivirus.
REFERENCE   1 (sites)
  AUTHORS   Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
             Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
  TITLE   Comparison of full-length sequences of interferon-sensitive and
             resistant hepatitis C virus type 1b. Sensitivity to interferon is
             conferred by amino acid substitutions in the NS5A region
  JOURNAL   J. Clin. Invest. 96 (1), 224-230 (1995)
  MEDLINE   95340824
  PUBMED   7542279
REFERENCE   2 (bases 1 to 9410)
  AUTHORS   Enomoto,N.
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 9410)
  AUTHORS   Enomoto,N.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-MAY-1995) Nobuyuki Enomoto, Tokyo Medical and Dental
             University, Second Department of Internal Medicine; 1-5-45 Yushima,
             Bunkyo-ku, Tokyo 113, Japan (E-mail:PXN04522@niftyserve.or.jp,
             Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
FEATURES   Location/Qualifiers
     source   1..9410
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="HCV-1b"
                     /db_xref="taxon:11103"
                     /clone="HCV-K1-S2"
                     /note="IFN-sensitive HCV from patient 2"

     CDS   330..9362

Qy   3361 CGGCTTTATATCGGGGCGTGGGTTCCCCTTGACTAATTCAAAAGGGCAGAACTGCCGGCTATCGCCGG 3420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8361 CGGCTTTATATCGGGGCGTGGGTTCCCCTGACTAACTCAAAAGGGCAGAACTGCCGGTTATCGCCGG 8420

Qy   3421 TGCCGGCGCCAGCGCGGCGGGTGCTAGCTGCGTAATACCCTCACATGTTACTTGAAG 3480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8421 TGCCGGCGCGAGCGCGGGGGTGCCGACCAGCTGCGTAATACCCTCACATGTTACTTGAAG 8480

Qy   3481 GCCTTCTGCGAGCCTCGCGAGCTCGAAAGCTCCAAGACTGCTTCGTTGTCCGTGCGGGGAGAC 3540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8481 GCCTTCTGCCAGCCTCGTCGAGTGCCAAGCTCGACAGACTCGTTGTCCGTGCGGGGGAGAC 8540

Qy   3541 GGCCCTTGTCGTTATCTGTGAGAGCGGGGAACCAGGAGGAGCGGGGAGCGGGGACCTAGGAGTC 3600
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8541 GACCTCGTCCGTTATCTGTGAAAGCGGGGAGACCAGGAGCGGGGAGCGGGGACCTAGGAGTC 8600

Qy   3601 TTCACGGAGGCTATCGACTAGCTACTTGTGCCCCCCGGGGAGGACCCGGGGAGGAGCCGGAATAC 3660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8601 TTCACGGAGGCTATCGACTAGCTACTTCTGCCCCCCGGGGGGACCCCCAAAACCAGAATAC 8660

Qy   3661 GACCTTGGACTTGATTAAACATCATGTCCTCCTCCCCCGGGGGCGACGATGCAGTCTGGCC 3720
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8661 GACTTGGGAGTTGATTAAACATCATGTCCTCCTCCCCAAATGTGTCCGTCGGGGCGCACATGGC 8720

Qy   3721 AAAAGGGTATTACTTACCTCACCGTGACCC 3749
            ||||||||||||||||||||||||||
Db   8721 AAAAGGGTGTTACTTACCTCACCGTGACCC 8749

     mat_peptide   330..902
                     /product="core protein"
     mat_peptide   903..1478
                     /product="E1"
     mat_peptide   1479..2756
                     /product="E2"
     mat_peptide   2757..3407
                     /product="NS2"
     mat_peptide   3408..5300
                     /product="NS3"
     mat_peptide   5301..5462
                     /product="NS4A"
     mat_peptide   5463..6245
                     /product="NS4B"
     mat_peptide   6246..7586
                     /product="NS5A"

     CDS   330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAA09076.1"
                     /db_xref="GI:1814089"
                     /translation="MSTNPKPQRTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGASRALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNASGVYHVTNDCSNASIVYEAAD
                     MIMHTPGCVPCVREANNSRCWVALTPTLAARNSSVPTTIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVV
                     SQLLRIPQAVMDMWAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGRTTVVGGAQ
                     GHTTQRLASLFTTGGAQORIQLVNTNGSWHINRTALNCNDSLNTGFLAALFYTHRFNSS
                     GCPERIASCRSIDAPAQGMWNFGQTPITYYGPISSDQRPYCWHYAPRPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRSGAPTYSWGENETDVLLLNNTRAPQGNWFGCTWMNGTGFTKTCGG
                     PPCNIGGVGNNTLTCPTDCFRRHPEATVTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PVLSTGLIHLHQNIVDVQYLYGVGSAAVSFVIKWEYILLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAWYIKGRLVPGAAYAFYGVWPLLL
                     LLALPPRAYAMDREMAASCGGAVFVGLALLTLSPHYKVVLARLIWWLQYFITRAEAC
                     LQVWIPPLNVRGGRDAIILLTCAIHPELIFSITKILLAILGPLMVLQAGITRMPYFVR
                     AQGLIRACMLVRKVAQGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVE
                     PVVFSDMETKIITWGADTANAACGDIILGLAVSARRGKEIFLGPADSLEGGWRLLAPIT
                     AYSQOTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCTNGVCWTVYHGASK
                     DGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPVEYAAQCYKVLVLNPSVAATL
                     GFGAYMSKAHGVDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNIGEIPFYGK
                     AIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPASGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RAGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVVQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTLLYRLGAVQNEVILTHPITKYIMACMSADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGKPAIIPDRELVYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWQALETFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPRAASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGDMPSTEDLVNLLPAILSPGALVVGVVCA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSRVLLRLPGVPFLSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNLKCPCQVPAPEFFTEKELDGVRLHR
                     YAPASKPLLRDEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAEAAKRRL
                     ARGSPPSLASSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGGNINTRVESEN
                     KVVILDSFEPLRAEEDEREVSLPAEILRKSRKFPPAMPIWARPDYNPPLLESWKDPDY
                     VPPVVHGCPLPPRKAPPIPPPRRKRTVVLTESTVSSALAELATKTFGSSESSSSDSGT
                     ATAPPDQPSNDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMS
                     YTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKARLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSRAINHI
                     RSVWKDLLEDTEPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDV
                     VSTLPQAVMGPSYGFQYSPGQRVEFLVNAWKSKKCPMGFAYDTRCFDSTVTESDIRV
                     EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
                     TCYLKATAACRAAKLQDCTMLVCGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPG
                     DPPQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
                     GNIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
                     GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRASVRAKLLSQGGRAICGKYL
                     FNWAVRTKLKLTPIPAASQLDLSGWFVAGYSGGDIYHSVSRARPRWFMWCLLLSVGV
                     GIYLLPNR"

```
mat_peptide      7587..9359
                 /product="NS5B"
misc_feature     1479..1559
                 /note="hypervariable region (HVR)"
misc_feature     6954..7073
                 /note="interferon sensitivity determining region (ISDR)"

ORIGIN

Query Match          88.3%;   Score 3312.2;  DB 14;  Length 9410;
Best Local Similarity   92.7%;   Pred. No. 0;
Matches 3476;   Conservative  0;  Mismatches 273;  Indels  0;  Gaps  0;
```

Qy   1 TGGGAGGGGGTCTTCACAGGCCCTCACCCACGTGGAGTGGCCCCACTTCCTTGTCCCAAACAAAG 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5001 TGGGAGGAGCCTCTTCTTTACAGGCCCTCACCCACCCACATAGGATGCCCACTTCCTTGTCCCAGACTAAG 5060

Qy   61 CAGGCCAGGAGGAGCAACAGTTCCCCTACCTTGGTGGCGTACCAGGCCTACTTGTGTGTCGGGCCTAGGGCC 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5061 CAGGCAGGAGAGACAACTTCCCCTACCTTGGTAGCATACCAACGCCTACCGGTGTGCGGCCAGGGCT 5120

Qy   121 CAGGCCCCACCTCCATCGGGGATCAAAATGTGGGAGGAAGTGTCTTCATACGGGTAAAAGCCTACT 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5121 CAGGCTCCACCTCCATCGGTGGGAATCAAAATGTGGGAAGGTGTTCATAAAGGCCTACG 5180

```
Db   6981  CGTCACGACTCCCCAGATCCGACCTCATCGAGCGCCAACCTCTCGTGGGGGGCAGAGATG  7040
Qy   2041  GGCGGGGACATTACCCGGTGGAGTCAGAGAACAAAGGTAGTAATCCTGGACTCTTTCGAC  2100
                ||||||| |||||||||||| |||||  || |||||||| || |||||||||||||||
Db   7041  GGGGGGGACATCACCCGGCGTGGAGTCAGAGAATAAGGTGGTGATTCTGGACTCTTTCGAG  7100
Qy   2101  CCGCTTCGAGCGCGAGGGAGGGAGGAAGTGTCCGTGCCGGGAGATCCTGGGGAAA  2160
             ||||| ||||||| |||||||||  ||  |||||||||||||||||||||| ||||
Db   7101  CCGCTTCGAGCGCGAGGGAGGGAGGAAGTATCCTCCGGGGGGAGGATCCTGGGGAAA  7160
Qy   2161  TCCAAGAAATTCCACCACGGATGCCCGCGCCGGGCCGGATTACAAACCCTCCGCTG  2220
             ||||||||||| ||| || |||||||| || ||||||||||| || ||||  |||
Db   7161  TCCAGGAAATTCCCCCGGCAATGCCCATATGGGGCACCGCCGGATTACAAACCCTCCACTG  7220
Qy   2221  CTGGAGTCCTGGAAGGCCCCGGACCTACGGTTCCCTCCAGTGGTACATGGGGTGCCCACTG  2280
             ||| |||||||||||||||| ||||| |||||||| |||||||||||||||||||  ||
Db   7221  ATAGAGTCCTGGAAGGGACCCGGACGTCCCTCCAGTGGTGCACGGGGTGCCCATTGCCA  7280
Qy   2281  CCTACTAAGACCCCCTCCTATACCACCTTCACGGAGGAGGACAGTTGTTCTGACAGAA  2340
             ||||||| ||||||| || ||||||||||||||||| || ||||||||| ||||||||
Db   7281  CCTACCAAGGCCCCTCCAATACCACCTCCACGGAGGAAGACAGGTTGTTCCTGACAGAG  7340
Qy   2341  TCCAACCGGTGTCTTCTTGCCCTGGGCGGAGCTTGCCACAGAGCTTCCGACACCG  2400
             |||| || ||||||| ||||||| || |||||||||||||||||||||||| ||||
Db   7341  TCCACTGTGTCTTCTTGCCTTGGGCGGAGCTCGCGCCAGCCTGGGCAGCTCCGAATCG  7400
Qy   2401  TCCGGCCGTCGACAGGGCACGGGCAACCGCTCGGCAACCTTCGGGACTCGGGCGGGGGA  2460
             ||||| || ||| |||| || ||  |  |  |||| |||| |||| |||||||||| ||
Db   7401  GCAGGATCCGACGTTGAGTGCGTACTCCTCCTCCACTGTAGCAGTTGGGGGGGGAC  7460
Qy   2461  GCAGGATCTGACGTTGAGTGTGTATTCCTCCATGCCCATGCCCCCCTTGAGGGGGAGGAGAC  2520
             ||||| |||||||||||| ||||| |||||| |||||| |||||||| |||||||||||||
Db   7461  GCAGGATCGGACGTTGAGTGTGTACTCCTCCATGCCCATGCCCCCCTTGAGGGGGAGGGAC  7520
Qy   2521  CCCGATCTCAGCGCGACGGGGGTCTTTGGGTCTACGGTGACTGGAGGAATTCTGTC  2580
             |||||||||||| |||||| |||||| |||||| |||||||||||||||| |||||
Db   7521  CCCGGATCTCAGCGACGGGGGTCTTTGGTCTACGGTGAGGGATGTCGTC  7580
Qy   2581  TGCTGCTCGGATGTCCTACACATGACAGGCCGCTCTGATCAGCCATGGCCTGCGGCGGGAA  2640
             ||||| || ||||||| |  |  |||||||||||||| |||| |||||||||||||||||
Db   7581  TGCTGCTCGGATGTCCTACACCATGGACAGGCCCTGATCACACCATGGCCTGCGGCGGGAA  7640
Qy   2641  AGCAAGCTGCCCATCCAAGTGCCTTGAGCAACTCTCTTTGCTCCACCAAACACATGGTTCTAC  2700
             |||||||| |||||| || |||||||||||||| || || |||||| |||||||| |||||
Db   7641  AGCAAGCTACCATCAAATGCCTTGAGCAACTCTTTTGCTCGCCCCACCCATAATATGGTTCTTAT  7700
Qy   2701  GCTACCACCATCCCGCAGCGCGAGCGAGAAGGTCACTCTTTTGACAGACTGCAA  2760
             |||||| ||||||| |||||||||||||| ||||||||||||| |||||||||||
Db   7701  GCCACAACATCCCGAGGGCGCAAGCGCCAAAGGTCACCTTTTGACAGACTGCAA  7760
Qy   2761  ATCCTTGGACGATCACTACACGGACGTGCTCAAGGAGATGAAGGCGAAGGCCGTCACAGTT  2820
             ||||| |||||| ||||||||||| ||||||||||||||||| ||||||||||| ||||||
Db   7761  GTCCTTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCCGTCACAGTT  7820
Qy   2821  AAGGCTAAGCTTCTATCAGTAGAGAAGAAAGCCTGCAAGGTCGACGGCGCCCCACATTCGGCCGAA  2880
             |||||| ||||||||||||| || |||||| |||||||||||||||||| ||| |||||
Db   7821  AAGGCCGACTCTTATCCGTGTTGAGGAGGCCTGTAAGCTGACGCCCCACATTCGGCCGAGA  7880
Qy   2881  TCTAAATTTGGCTATTGGGGCAAAGGAGCGTCCGGGAACCTATCCAGCAGGCAAGGCCATTAACCAC  2940
             |||| ||||||||||| |||||| ||||||| |||||||| |||||| ||  |||||  |||| |||
Db   7881  TCCAAATTTGGCTATGGGGCAAAAGAGCGTCGGGGACGTCCTATCCAGCGCAAGCCATTAACCAC  7940
Qy   2941  ATCCGGCTTCCGTGTGGGAGGACTGTTGTTGGGAAGACACTGAAACACCAATTGACACCACCATC  3000
             ||||||  |||| || ||| ||| |||||||||||| ||| || ||||||| |||||||||
Db   7941  ATCCGTCCCGTGTCGAAAGACACTGAGACACTGAAAGCACTGTTTGACACCATC  8000
Qy   3001  ATGGCAAAAAATGAGGTTTTTCGCTCCAACCAGAGACGAGAGGAGGAGGCCAAGCCAGCTCGC  3060
             |||||||||||||| |||||| ||||||||||||||| |||||||||| ||||||||||||
Db   8001  ATGGCAAAAAATGAGGTTTTTCGCTCCAACCAGAGACGAGAGGAGGCCAAGCCAGCTCGC  8060
Qy   3061  CTTTATCGTGTGTTCCCAGATTTGGGGGTTGCGGTGTGTGTGGGAGAAAAATGGCCCCTTTACGACGTG  3120
             ||| |||||||||||||||||||| |||||| |||||| ||||||||||||||  ||||||||||||
Db   8061  CTTATCGTGTTCCCAGACTTGGGGGTCCGGGTGTGTGAGAAGAAAGTGGGCCCTTTACGACGTG  8120
```

```
Qy   3121  GTCTCCACCTCCCTCAGGCTGTGTGAGATGGGGGTCCTTCCTGCGGCTTCCTCGTGGGGATTCCAGTTT  3180
            |||||||||    |||  ||||||  ||||||| |   |  | ||||||| |||||    |||||||
Db   8121  GTCTCCACCCCTTCCTTCAGGGCCGTGATGGGGGGCCCCGTCATACGGACATTCCAATACTCTCCTGGA  8180
Qy   3181  CAGGCGATTGGAGTTCCTCGTGTGAACGAGCCCGTGAAAAGAGAGACCCCTATGGCGCCGGATTCCAATACTCTCCTGGA  3240
            |||||||||||| ||| |||   |  |||||||||||| |||||||||||| ||||| |||||||| |
Db   8181  CAGCCGGGTCGGAATTCCTGGTTGGAATGGGGCCCCTGATGGGCGCCCTATGGGTTTTGCA  8240
Qy   3241  TATGACACCCGTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTTCGACTGTTGTAGAGGAGTCA  3300
            |||||| ||||| ||||||||  |||||| ||||| |||  ||| || |||  ||||||| |||||||
Db   8241  TATGACACCCGCTGTTTGACTTCAACGGTCACTCGAGTACATCCGTGTTTCAGTTGTTGAGGAGTCA  8300
Qy   3301  ATTTATCAAGTGTTGTGACTTGGCCCCCCGAAGGCCAGAGGCCATAAGGTCGCTCACAGAG  3360
            ||  ||||||| ||||| ||||||||||| |||| |||| |||| || ||||||||||||
Db   8301  ATATACCAAATGTTGTGACTTGGCCCCCCGGAGGACAGCAGGCCCATAAGGTGGCTCACAGAG  8360
Qy   3361  CGGCGTTTATATCGGGGGTCCCGTGTGAACGACCCGTAATAATTCAAAAGGGGGGTATCGCGCCCG  3420
            |||| |||||| |||||| ||||||||| ||   | |  || |||| | |||||| |||||| ||
Db   8361  CGGCTTTATATCGGGGGTCCCCGTGTGAACTGACTAATCAAAAGGGGGTGATCGCGCCCG  8420
Qy   3421  TGCCGCGCGGAGCGGGGTCGGTGCGCTGGTGCGTGAGAGTGGTAATACCCTCACATGTTACTTACTTGAAG  3480
            ||||||||||||||| || |||||| || || ||||||||| ||||||||||| | ||||||||| |||
Db   8421  TGCCGCGCGGAGCGGGGTCGGTGCGCTGGTGCGTGAGAGTGGTAATACCCTCGCATGTTACTTGAAG  8480
Qy   3481  GCCTTGCTCGGTTGATAACATACTTAGGTACTTCGTGGAGGCGGCGGGACCCAGGACTGGTAATACCCCTCACAGTC  3540
            |||| || |||||||||||| ||||  |||||||| ||||||||| || || |||||||| |||||||||||||
Db   8481  GCCACTGCGGTCTGATAACATAGTCAGGTACTTCGTGGAGGCGTCGGGACCCAGGACTCGAGCCCAGGAC  8540
Qy   3541  GGGCTTGTCGTTATCTGTTATACATGTGCTCCTCCTCCTTCCATTGTGTCCGGGGTCTGGCCATCTGGC  3600
            ||||  || ||||| ||| ||||||||||| ||||| |||||||| ||  ||||  |||||||||||||
Db   8541  GGGCCTGTCGTTATCTGTTATACATGTGCTCCTCCTCCTTCCATTGTGTCGGGGGTCTGGCCATCTGGC  8600
Qy   3601  TTCACGGAGGCTATGACTAGGTACTCTGACGGGGGAACCCCCAGCCCCAACCAGAATAC  3660
            |||||||||| || |||| | |||||  |  ||||| || |  | |  || ||||||||
Db   8601  TTCACGGAGGGCTATGATAGGTACTCTGTGAAAGGCCGGGAACCGCCAAGCCTACGAGTC  8660
Qy   3661  GACCCTTGGAGTTGATAACGTCCTCAACATGCTTCTTCCAATGTCTGGCCGGCATCCTGGC  3720
            ||| || ||||||||||||| | ||||| |||| || || |||||| ||| |||||||||
Db   8661  GACTTGGAGTTGATAACATCATCCTTCCAATTGCTCTCCCCAATGTCGTCGGCGATGCATCCGGC  8720
Qy   3721  AAAAGGGTATACTACCTCACCCGTGACCC  3749
            ||||||||| || |||||||||||||||
Db   8721  AAAAGGGTGTACTACCTCACCCGTGACCC  8749
```

RESULT 6
```
HPCK1R2
LOCUS        HPCK1R2        9410 bp        RNA        linear        VRL 10-FEB-1999
DEFINITION   Hepatitis C virus (strain HCV-1b, clone HCV-K1-R2), complete genome
             sequence.
ACCESSION    D50481
VERSION      D50481.1  GI:1030705
KEYWORDS     polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
             NS5B; envelope protein; non-structural protein;
             interferon-sensitive; interferon-resistant; IFN-sensitive;
             IFN-resistant; ISDR; interferon sensitivity determining region;
             HVR; hypervariable region.
SOURCE       Hepatitis C virus
  ORGANISM   Hepatitis C virus
             Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
             Hepacivirus.
REFERENCE    1  (sites)
  AUTHORS    Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
             Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
  TITLE      Comparison of full-length sequences of interferon-sensitive and
             resistant hepatitis C virus 1b. Sensitivity to interferon is
             conferred by amino acid substitutions in the NS5A region
  JOURNAL    J. Clin. Invest. 96 (1), 224-230 (1995)
  MEDLINE    95340824
  PUBMED     7542279
REFERENCE    2  (bases 1 to 9410)
  AUTHORS    Enomoto,N.
```

```
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 9410)
AUTHORS     Enomoto,N.
TITLE       Direct Submission
JOURNAL     Submitted (08-MAY-1995) Nobuyuki Enomoto, Tokyo Medical and Dental
            University, Second Department of Internal Medicine; 1-5-45 Yushima,
            Bunkyo-ku, Tokyo 113, Japan (E-mail:PXN045222@niftyserve.or.jp,
            Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
FEATURES             Location/Qualifiers
     source          1..9410
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="HCV-1b"
                     /clone="HCV-K1-R2"
                     /db_xref="taxon:11103"
                     /note="IFN-resistant HCV from patient 2"
     CDS             330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAA09072.1"
                     /db_xref="GI:1814086"
                     /translation="MSTNPKPQRKTKNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARQPEGRANAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAAD
                     MIMHTPGCVPCVREANSSRCWVALTPTLAARNSSVPTTIRRHIDVDLLVGAAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVV
                     SQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGHTYVSGGAK
                     SHTTQGLVSLFRPGAQQKIQLVNTNGSWHINRTALNCNDSLNTGFLAALFYTHRFNSS
                     GCPERIASCRSIDAFAQGWGPITYAEPGSSDQRPYCWHYAPRPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRSGAPTYSWGENETDVLLLNNTRPPQGNWFGCTWMNGTGFTKTCG
                     PPCNIGGVGNNTLTCPTDCFRKHPEATVTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQILPCSFTTL
                     PVLSTGLIHLHQNIVDVQYLYGVGSAVVSFVIKWEYILLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVVLNAASVAGAHGFLSFLVFFCAAWYIKGKLVPGAAYAFYGVWPLLL
                     LLLALPPRAYAMDREMAASCGGAVFVGLALLTLSPHYKVVARLIWWLQYFITRAEAC
                     LQVIFPLNVRGGRDAILILLTCALHPELIFSITKILLVAILGPLMRQGYLGAITKMPYFVR
                     AQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTLPQDWAHTGLRDLAVAVE
                     PVVFSDMETKIITWGADTAACGDIILGLAVSARRGKEIFLGPADSLEGSGWRLLAPIT
                     AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYHGAGSK
                     TLAGPKGPITQMYTNVDQDLVGWQAPSGARSLTPCTCGSSDLYLVIRHADVIPVRRRG
                     DGRSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGVDPNIRTGVRTITTGSKFLADGGCSGGAYDIIICDECHS
                     TDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
                     AIPIEITIKGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPASGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTVPQDAVSRSQRGRTGRG
                     RAGIYRFVTPGERPSGMPDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVNQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGDNPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPPLHGPPLLPLYRLGAVQNEVILTHPITKYIMACMSADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWQALETFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGDMPSTEDLVNLLPAILSPGALVVGVVCAA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLSDFKTWLQSRVLPRLPGVPFLSCQRGYK
                     GVWRGDIMQTTCPCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNLKCPCQVPAPEFFKELDGVRLHR
                     YAPASKPLLRDEVTFQVGLNQYVVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRL
                     ARGSPPSLASSSASQLSAPSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVVILDSFEPLRAEEDEREVSLPAEILRKSRKFPRAMPIWARPDYNPPLIESWKDDY
                     VPPVVHGCPLPPTKAPPIPPPRRKRTVVLTESTVSSALAELATKTFGSSESSAADSGT
                     ATAPPDQPSSDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMS
                     YTWTGALITPCAAEEKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKARLLSVEEACKLTPPHSARSKFGYGAKDVRNLSRAINHI
                     RSVWKDLLEDETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
                     VVSTLPQAVMGASYGFQYSPGQRVEFLVNAWKSKKCPGEPAYLTDHCSTVTENDIRV
                     EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
                     TCYLKATAACRAAKLQDCTMLVCGDDLVICESAGTQEDAASLRVFTEAMTRYSAPPG
                     DPPQPEYDLELITSCSSNVSVAHDAGSKRVYLITRDPTTPLARAAWETARHTPVNSWL
                     GNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
                     GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGGRAAICGKYL
                     FNWAVRTKLKLTPIPAASQLDLSGWFVAGYSGGDIYHSVSRARPRWFMWCLLLLSVGV
                     GIYLLPNR"
     mat_peptide     330..902
                     /product="core protein"
     mat_peptide     903..1478
                     /product="E1"
     mat_peptide     1479..2756
                     /product="E2"
     mat_peptide     2757..3407
                     /product="NS2"
     mat_peptide     3408..5300
                     /product="NS3"
     mat_peptide     5301..5462
                     /product="NS4A"
     mat_peptide     5463..6245
                     /product="NS4B"
     mat_peptide     6246..7586
                     /product="NS5A"
     mat_peptide     7587..9359
                     /product="NS5B"
     misc_feature    757..9359
                     /note="hypervariable region (HVR)"
     misc_feature    1479..1559
     misc_feature    6954..7073
                     /note="interferon sensitivity determining region (ISDR)"
ORIGIN
     Query Match
     Best Local Similarity   88.1%;   Score 3302.6;   DB 14;   Length 9410;
     Matches 3470;  Conservative  92.6%;  Pred. No. 0;  Mismatches 279;  Indels  0;  Gaps  0;

Qy        1  TGGGAGGGGCGTCTTCACAGGGCCTCACCCACGTGGATGCCCCAACACTTCCTGTCCCAAAACAAAAG        60
             |||||||||  |||||||||||| |||||||||||||||||||||| |||||||||||||||||||
Db     5001  TGGGAGGAGGCGTCTTCTTACAGGGCCTCACTCACATAGATGCCCATTTCCTTGTCCGGTCCAGGGGCC      5060

Qy       61  CAGGGCAGGAGGACAGACACTTCCCTCACCTGGTCTGGCGTACCAGGCCACTTGTGTTCCGGCTAGGGGC      120
             |||||||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||| ||
Db     5061  CAGGCGAGGAGGACAACTTCCCATACCTGGTAGCATACCAGGCTACGGTGTGTGCGACTAGGGCT         5120

Qy      121  CAGGGCCCCACCTTCCATCATGGGATCAAAATGTGGAGAGTGTCTCATACGGCCTAAAGCCTTACT       180
             |||||||||||||||||||||||||||||| |||||||||||||||||| |||||||||||
Db     5121  CAGGCTCCACCTCCATCGTGGGATCAAAATGTGGAGAGTGTCTCATACGGCTAAAGCCTACG         5180

Qy      181  CTGCGCGGGGCCCAACACCCTTGCTGTATAGGCTGGGAGGCCGGTCCAAAAACGAGGTCACCCTC      240
             |||| ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     5181  CTGCATGGGGCCCAAACACCCTGCTGTATAGGCTGGGAGGCCGTCCAAAAACGAGGTCATCCTC      5240

Qy      241  ACACACCCCATAAACCAAATTCATCATGCATGGGAGTGTCAGCCGACTAGGCGAGGTCTTCACG      300
             |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     5241  ACACACCCCATPAACCAAATACATCATGGCATGGGAGTGTCGGCCGTCACCTGACGGTCGTTACG       5300

Qy      301  AGCACCTGGGGTGCTGGTGGGGGGTCCTTGGCCAGGTCTTGCCGTATTGCTTGGTGACAACA      360
             |||||||||||||||||||||||||||| |||||||||||||||||| |||||||||||
Db     5301  AGCACCTGGGGTGCTAGTAGGCCGCGGCGGTCCTGGCTCTTGGCCGTATTGCCTGACAAACA        5360

Qy      361  GGCACGTGGTCATTGGGTTAGGATCATCTTGTTCCGGGCCGGCCGCCTATTGTTCCCGAC       420
             ||| ||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     5361  GGCACGGTAGTCATTGTGGGCAGGATCATCTTGTCCGGACCCGGCCATCATTCCCGAC       5420

Qy      421  AGGGAAGTCCTCTCACCAGGAGTTCGATGAGATGGAAGAGTGCGCGGTCGCACCTCCCTTAC      480
             |||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     5421  AGGGAGGTTCTCTACCGGGAGTTCGATGAGATGGAAGAGTGCGCGTCACACCTTCCTTAC       5480

Qy      481  ATCGGAGCAGGGAAATGCAGTCGCGCTGGGGCCGAGCAGTTCAAGCAAAAAGGCCTCGGGTTGCTGCAG  540
             ||||| ||||||||||||||||||||||||| ||||||||||||||||||||||| ||||||||
Db     5481  ATCGGAGCAGGGAATCGCAGCTCGGCCGAGCAATTCAAGCAGAAAAGCCTCGGGTTGCTGCAG       5540

Qy      541  ACAGCCACCAAGCAAGCGGAGGCCGCCGCTGTCCCGGTGGTGGAAAGCAAGTGGCAGGCCCTT       600
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||| |||||
Db     5541  ACGGCCACCAAGCAGGCGGAGGCCGCCGCCCTGGTGGAGTCCAAGTGGCAAGCCCTT       5600

Qy      601  GAGACCTTCTGGGCGGAAAACACAATGTGGAACTTCATCAGCGGAATACAGTACTTAGCAGGC       660
             ||||||||| ||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     5601  GAGACCTTTTGGGCCGAAAACACAATGTGGAATTTCATCAGCGGAATACTTAGCAGGC        5660
```

```
Db  7821 AAGGCCAGACTCTTTATCCGTAGGGGGGAGGAGGCCTGTAAGCTGACGTGACGCCCCCAGA 7880

Qy  2881 TCTAAATTTGGCTATGGGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 2940

Db  7881 TCCAAATTTGGCTATGGGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 7940

Qy  2941 ATCCGGCTTCCGTGGGGAGGACTTGTTGGGGAGACTTGTTGGGGAGACACACAATTGACACACCATC 3000

Db  7941 ATCCGGCTCCGTGGGGAGGACTTGTTGGGGAGACTTGTTGGGGAGACACACAATTGACACACCATC 8000

Qy  3001 ATGGCAAAAAATGAGGTTTTTCTGCCTCCAACCAGACACTGAGACCAATTGACACCACTACGC 3060

Db  8001 ATGGCAAAAAATGAGGTTTTTCTGCCTCCAACCAGAGAGGAGGCCGCAAGCCAGCTCGC 8060

Qy  3061 CTTATCGTGTGTTCCAGACTTGGGGGGTGCCGTGGTGTGTGCGAGAAAATGGCCCTCTATGACGTG 3120

Db  8061 CTTATCGTGTTCCCAGATTTGGGGGGTTCGTTGTGTTCGGGAGGAAATGGCCCTTTACGACGTG 8120

Qy  3121 GTCTCCACCCTCCCTCAGCTCAGCGCTTGTGATGGGTTCTCTCGTTAGCGGATTCCAGATTCTCTCCTGGA 3180

Db  8121 GTCTCCACCCTTCCTCAGGCGCTGTGATGGGGGCCTCATACGGTACTCTCCTGGA 8180

Qy  3181 CAGGGGGTCGGAGTTCCTGGGTGAACCCCTGGAAATCAAAGAAAGAACCCTATGGGGCTTTGCA 3240

Db  8181 CAGGGGGTCGGAGTTCCTGTGAATGCCTCGGAAATCAAAGAAAGAATGCCCCATGGGGCTTGCGCA 8240

Qy  3241 TATGACACCCGCTGTTTGACTCAACAGTCACTGAGAATGACATCCGTGTTGTTGAGAGGAGTCA 3300

Db  8241 TATGACACCCGCTGTTTTGACTCAACAGGTCACTGAGAATGACATCCGTGTTGAGAGGAGTCA 8300

Qy  3301 ATTTATCAAATGTTGTTGACTTGGCTGGCCCCGAGAGCAGCCAAGGTTGTCGCTCACAGAG 3360

Db  8301 ATATACCAATGTTGTTGACTTGGCTGGCCCCGAGAGCAGCCACAGCACAGGGTCGCTCACAGAG 8360

Qy  3361 CGGCTTTATATCGGGGGTCCCCTGACTAATTCAAAGGGCCAGAACTGGGCTATCGCCGGG 3420

Db  8361 CGGCTTTATATCGGGGGTCCCCTGACTAATTCAAAAGGGCCAGAACTGCGTCGGGTTATCGCCGG 8420

Qy  3421 TGCGCGGCCAGCGCGGTGACGTAGCTGCGTGCTAATACCCTCACATGTTACTTGAAG 3480

Db  8421 TGCGCGCAAGCGGGTGCTGTGCTAATACCCTCACATGTTACTTGAAG 8480

Qy  3481 GCTCTCGAGCGTCTGCTCGAGCTGCAAAGCTCCACAGACTGCTCGTCGTGTGTGGGGGAC 3540

Db  8481 GCCACTGCGCCCGTCTGCAAGCTCCAGGACTCTCGTTGTGTCGGGAGGAC 8540

Qy  3541 GGCCTTGTCGTTATCGTGTGAGAGCGCGGCGGAGCGGGAGCCTACGAGTC 3600

Db  8541 GACCTTGTCGTTATCGTGTGAAAGCGGCGGGAAACCCAAGCCTACGAGTC 8600

Qy  3601 TTCACGGAGGCTATGACTAGGTACTCTGCCCCGCCCCCCAGGGGGACCCCCCAGAATAC 3660

Db  8601 TTCACGGAGGGCTATGACTAGGTACTCTGCCCCCCAACCAGAATAC 8660

Qy  3661 GACTTGGAGTTGATTAAACATCATCTGCTCCTCCCAAATGTGTGTCGGGGCACCGATGCATCTCGGC 3720

Db  8661 GACTTGGAGTTGATAAACATCCTCCCAATGTGTCCAACCCACCATCCGGC 8720

Qy  3721 AAAAGGGGTATACTTCACTCCACCCGTGACCC 3749

Db  8721 AAAAGGGGTACTACCTCACCCGTGACCC 8749
```

```
VWKDLLEDTETPIDTTIMAKNEIFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVV
STLPQAVMGSSYGPQYSPSGQRVEFLVDAWKSKKCPMGFAYDTRCFPDSTVTESDIRVEE
SIYQCCDLAPEARQAIRSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTC
YLKASAACRAAKLRDCTMLVCGDDLVICESAGTQEDAASLRVFTEAMTRYSAPPGDP
PQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARSTPVNSWLGN
IIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGL
SAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYLFN
WAVRTKLKLTPIPAASQLDLSNWFVAGYSGGDIYHSLSRARPRWFMLCLILLSVGVGI
YLLPNR"
```

ORIGIN
```
Query Match     87.9%;  Score 3297.8;  DB 14;  Length 9373;
Best Local Similarity   92.5%;  Pred. No. 0;
Matches 3467;  Conservative  0;  Mismatches 282;  Indels  0;  Gaps  0;
```

```
Qy     1 TGGGAGGGGCGTCTTCACAGGCCTCACCCACGTGGATGCCCACTTCCTTGTCCCAAACAAAG 60
         |||||||| ||||||||||||||||||||||||    |||||||  ||||||||||| ||
Db  4995 TGGGAGGAGCCTCTTCACAGGCCTCACCCACATAGAGACGCCCACTTCTTTGTCCCAGACTAAA 5054

Qy    61 CAGGCCAGGGAGGAGCAACTTCCCCTCACCTGGTGGGGGTACCAGGCTACTGGTGTGCGCTAGGGCC 120
         ||||||||| || |||||||||||||||||||||||||||||  |||||| ||||||||||||
Db  5055 CAACCAGGAGATAACTTCCCCTACCTGGTGGGCATACCAAGCCTACAGTGTGCGCCAGGGCT 5114

Qy   121 CAGGCCCCACCTCACCTCCATCATGGGATCAAAATGTGGAAGTGTTCTATACGGGCTAAAGCCTACT 180
         |||||||||||  |||||||| ||||||||| ||||||||||||||||||||||  ||||||
Db  5115 CAGGCTCCACCCACCATCGTGGGGATCAAAATGTGGAAGTCGCCTCATACGGGCTAAAGCCTACG 5174

Qy   181 CTGCGGCGGCCAAACACCCTTGCTGTGTATAGGCTGGGGAGCCGTCCCAAAAACGAGGTCACCCTC 240
         ||||||| ||||||||||| ||||||||||||||||||||||||   |||||| ||||||||||
Db  5175 CTACACGGGCCAACGCCCTGCTGTGTATAGGCTAGGAGCCGTCCCAAAATGAGGTCACCCTC 5234

Qy   241 ACACACCCCCATAACCAAAATTCATCATGGCCATGCATGCATGTCAGCCGACCTGGAGGTGTCACG 300
         |||||||||  ||||| |||||||||||||||||| ||||||||||||||||||||||||||||
Db  5235 ACACACCCCATAACCAAAATACATCATGGCATGCATGCATGTCGCCTGACCTGGAGGTCGTCACG 5294

Qy   301 AGCACCTGGGGTGCTGTGGTGGGCGGGGGTTCCTGCTCTGCGCTATTGTTTGCCTGACAACA 360
         ||||||||||||||||| |||||||||||||||||||||| ||||| || |||||||||
Db  5295 AGCACCTGGGGTGCTAGTAGGTGGGCGGGGGTTCTTGCCGCTCTGCCGTATTGCCTGACAACA 5354

Qy   361 GGCAGCGTGGTCATTGTGGGTAGGATCATCTTTGTCCGGGCGAGCCGTCCGGAC 420
         ||||||||||||||| || |||||| |||||||||||||  |||| ||||||
Db  5355 GGCAGCGTGGTCATTGTGGGCAGAATCATCTTTGTCGGGGCAGCCGTCCGGAC 5414

Qy   421 AGGGGAAGTCCTCTACCCAGGAGTTCGATGAGAGTGGCGGCGTCGACCTCCCTTAC 480
         |||| ||||||||||||||||||||||||||| ||| || ||||||||||| ||
Db  5415 AGGGAAGTTCTCTACCCGGGAGTTCGATGAGAGAGCGGCGTCGACCCTCCCTAC 5474

Qy   481 ATCGAGCCAGGGGAATGCGCACTCGCCGAGCAGTTCAAGCAAAAAGCGCTCGGGGTTGCGTGCAG 540
         ||||| ||||| |||| |||||| || |||||||||||| ||||| |||||||| |||||||
Db  5475 ATCGAGCAGGGGGATGCGCCCCAGCAATTCAAGGACAGGCGCCTCGGGGTTGCTGACA 5534

Qy   541 ACAGCCACCAAGCAAGCGAGGAGCGGCCCCTGCTCCCGGTCAAGTGTGGCGAGCCCTT 600
         |||||||||||||| | ||||| |||||||||||| |||||||||||||||||||
Db  5535 ACAGCCACCAAGCAACGGGAGGTTGCGGCTCCCGGTCAAGTGTGGCGAGCCCTT 5594

Qy   601 GAGAACCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTGACTTAGCAGGC 660
         |||| |||||||||| ||||||| |||||||||||||||| || ||||||| |||||||||
Db  5595 GAGGCCTTTTGGGCGAAAGCACATGTGGAACTTCATCAGCGGGGATCACAGTGAGTTAGCAGGC 5654

Qy   661 TTGTCCCACTCTGCCTGGGGAATCCGGGATTGCATCACTACTGATGGGCGTTCACAGCCTCTGTC 720
         ||||||||||||| ||||||||||||||| |||||| ||||| |||| || |||||||
Db  5655 TTGTCCACCTTACCTGGAAACCCCGGGAATCCGGGCAATTCACACGGTTGGG 5714

Qy   721 ACTAGCCCGTCACCCCAAATCTACCCTCTCCTGGGGCGGTCGGGGTAGCC 780
         |||||| || |||||||| ||||| ||||||||| ||| ||||||||
Db  5715 ACCAGCCCGTCACCCAAGTAACCTCTGTTTAACATCTTGGGAGGATGGGGTGGCC 5774

Qy   781 GCCCAAACTCGGCTCCCCCAGTGGCGTCGCCTCAGCTTTCGTCAGGATGGTGG 840
         |||||| |||||||||||||||  |||  ||||||||||||||||||
Db  5775 GCCCAAACTCGCTCCCCCAGAGGCCGCTTCCTTGGCTTTCGTAGGCTCTGGTGCG 5834

Qy   841 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGGCTTGTTGTGGGACATCTTGGGCGGGCTATGGGAGCA 900
         |||||||| |||||| |||||| |||||||||||||||||||||||||| ||||||| |||||
Db  5835 GCTGTTGGCCAGCATAGGCCTTGGGAAGGTGCTTGTTGTGGGACATTCTGGGCGGGCTATGGGGCA 5894
```

```
Db  5835 GCTGTTGGCCAGCATAGGCCTTGGGAAGGTGCTTGTTGTGGGACATTCTGGGCGGGCTATGGGGCA 5894

Qy   901 GGAGTGGCAGGGCGGCGCTCGTGGCCCTTTAAGGTCATGAGAGCGGGGAAATGCCCTCCACCGAG 960
         ||||||||||| ||| |||||||||||||||||||||||||| ||||||||||||||||| |||
Db  5895 GGGGTGGCGGCGGGGCGCACTCGTGGCCTTTAAGGTCATGAGTGGCGAGATGCCCTCCACTGAG 5954

Qy   961 GACCTGGTTAACTTACTCCCTGCCATCCTTCTCTCCTTGTGCGTCCCTGGCTCGGTCGCGTG 1020
         |||||||||||||||||||||||||||||| ||||||||| ||||||||| || ||||||
Db  5955 GACCTGGTTAACTTACTCCCTGCCATCCTCTCTCCTTGTGGTCGCCTGGCTGGGCGTCGGTG 6014

Qy  1021 TGCGCAGCGGATACTGCCGTCCGCACGTGGGTCCAGGTGCAGTGCAGTGGATGGAAC 1080
         ||||||||  || |||| |||||||||||||| ||||| || || |||| |||||
Db  6015 TGCGCAGCCAATAACTGCGCCCGCACGTGGGGCCCAGGAGCGGCTGTGCAGGAGATGAAC 6074

Qy  1081 CGGGCTGGATAGGCGTTCGCCCTCCGCGGGGGGTTAACCAGCACTGTTGTGGCCAGAG 1140
         ||||||||||||| |||||| ||||||||||||||| ||| ||| || ||||||||
Db  6075 CGGCTGATTAGGCATTCGGCTTCGGCGGGGGTTAACCAGCACTATTGTTGCCGAG 6134

Qy  1141 AGCGACGACGGGCAGACGTGTCACTCAGATCCTCTCCGGACCTTACTATCACCAACTGTTG 1200
         ||||| || || |||||||||| |||| ||||||||||||| |||| ||||| |||||
Db  6135 AGCGGACGCTACGGGCGGGGTCACCAAAATCCTCTCCCAGCCTTACCATCTCAGTTGCTG 6194

Qy  1201 AAGAGGCTCCACCAGTGGACTTGGCATGGACTGCTCCACGCCCTGTCTGCTCCGGGGCTA 1260
         ||||| |||||||| |||| |||| || |||||||| |||||||| |||||||||||
Db  6195 AAGAGGCTCCATCAGTGGATCAAATGGACTGCTCCACGCCATGCCTTCCGGCGGTCCGGGGA 6254

Qy  1261 AGGGGATGTTTGGGACTGGATAGTGCACAGTTTGCTTGACTTCAAGGACCTGGGCTTCAGTCC 1320
         ||||||||||||| |||||||||||| ||||| |||||| ||||||||||||| |||||
Db  6255 AGGGATGTTTGGGACTGGGATATGCACAGTGTTGACTGACTTCAACGACCTTCCGTCC 6314

Qy  1321 AAGCTCCTGCCGCGAGTTACCGGGAGAGTCCCCTTTTTCTCATGGCCAACGCGGGTACAAGGG 1380
         |||||||| ||||||||||| ||||| ||||||||||| |||||||| ||| |||||||||
Db  6315 AAGCTCCTGCCCACCGGTTACGGGGAGAGTCCCTTTTTCTCAATGCCGCAACGCGGGGTACAAGGGA 6374

Qy  1381 GTCTGCGCGGGGGGAGACGGTCACCATGCAGCCACCTGCTCATGTGAGACACAGATCACCGGA 1440
         |||||||||||||||||||| || ||| ||||||| || ||| ||| |||||||||||
Db  6375 GTCTGCGCGGGGAGGGCATCATGCAGCCCCATGTGGTCGGCAAATTACCGGA 6434

Qy  1441 CATGTCAAAAACGGTTCCATGAGGATCGGTGGGGGCCTAAAATGCCCGTGCCAGTTCCA 1500
         |||||||||| |||||||||||||||| | ||||||||||| ||||| |||||||
Db  6435 CATGTCAAGAACGGTTCCATGAGGATGACACTGAGAGCAAACTGGGCCAT 6494

Qy  1501 GGAACAATTCCCATCAAACGGATAACACACCAGCGGGCCCCCTGCAGGTACGGCGTGGC 1560
         ||| |||||||| || ||||||||||| || || |||||| |||| ||||| || |||
Db  6495 GGAACAATTCCCAGCGGATAACACACCACCACGGGCCCCTGCAGGTACGGCGTGGC 6554

Qy  1561 TATTTCAGGGCGGCTGTTGGCGCGGCTGTGAGGAGTACGTTGGAGTTTACGGCGGG 1620
         |||||||| |||||||| || || |||| ||||| |||| |||| |||||||||
Db  6555 TATTCCAGGGCGCTATGGCCGTTGCTGTCTGAGGAGTTACGGCGGG 6614

Qy  1621 GATTTCCACTACGTGACGAGCCATGACCACTGACAAGCGTAAAATGCCCGTGCCAGTTCCA 1680
         ||||||||||||||| ||||| |||||||||||| |||| ||| |||||||||||
Db  6615 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTACCATGCCGGTTCCG 6674

Qy  1681 GCCCCCGGAATTCTTCACAGAAGTTGGATGGGGTGCGGCGGGTACGGTTCCGGTGC 1740
         ||||||||||||||||| |||||| ||||| ||| ||||||| |||| ||| |||
Db  6675 GCCCCCGGAATTCTTTCACAGAGAAGTGGATGGGGTGCGCCGGGTACGGTTCCG 6734

Qy  1741 AAACCTCTCCTACGGCGAGTCACATTCACAGCCTGGTTGGG 1800
         |||||||||| |||||||| ||| ||||| ||||||||||
Db  6735 AAACCTCTTCTACGGGGAGTCACAATATCCATGGTTGGG 6794

Qy  1801 TCGCAGCTCCCATGCGGAGCCCGAACCGGATGTAGCAGTAGTGCTCACTTCCATGCCGGAC 1860
         ||||||||||||||||||||||  ||||| ||||||||||| |||||||||||||| |||
Db  6795 TCACAGCTCCCTTGTGAGCCCGGACAGTGACGTCTCACTTCCATGCTCACCGGAC 6854

Qy  1861 CCCTCCCACATCACACAGAGAGGCGGCTTAAGCGGCAGGCAGCTGGCCGGCGTGG 1920
         ||||| || ||| |||||||||| || |||||||||||| |||| |||||||
Db  6855 CCCTCCCATATTACACGAGAGGGGGCTAGGCGGCAGGGGCTCTGGGCGTCGG 6914

Qy  1921 TTGGCCAGGCTCTTCAGCTAGCCAGTTGTCTGCGCGCCCCTCTTGGGACATAGCACTACC 1980
         ||||||||||| ||||||||||||||| || |||||| || |||||| |||||||||
Db  6915 TTGGGCCAGCTCTTCAGCTAGCCAGTTGTTGTCGCGGCGCCCTCTCTGGGACGGACATACATTACC 6974
```

```
Qy   1981  CAAAATGACTTCCCAGACGCTGACCTCATCGTCGAGGGCGCCAACCTTCCTGTGGGCGGCGGCATGACGTG  2040
           ||||| || ||||| || ||| |||||||||||| ||||| |||||||||||||||||||||     |||
Db   6975  CATCATGACTCCCCGGCACGCTGACCTCATCGAGGGGCCAACCTTCCTGTGGGCGGCAGCAGATG        7034

Qy   2041  GGCGGGGACATTACCCGCGGTGGAGTCAGAGAACAAGGTAGTAGTAATCCTGGACTCTTTGAC          2100
           ||||||||  || ||||||||||||  | |||| ||||| || | ||||| |||||||| |||
Db   7035  GGCGGAAACATCACCCGCGGTGGGGTCAGAGAATAAGGTAGTAGTAATTCTGGACTCTTTGAC          7094

Qy   2101  CCGCTCCGAGCGGAGGAGGAGTGACGGGGAAGTGTCCTGTCCCGGGGAA                        2160
           |||| ||||| |||||||||||| ||||||||||| || |||||||||
Db   7095  CCGCTCCGACCGGAGGAGGATGAGAGGAGGAGGGGTGTCCGTCCCGGGGAAA                     7154

Qy   2161  TCCAAGGAAATTCCCACCAGCGGATGGCCACCGGGGATTACAACCCTCCGCGTG                   2220
           ||||| ||||| |||||||||| |||||||||||||||||||||||||||||| |
Db   7155  TCCAGGAAGTTTCCCGCCAGCGGATGGCCATGGGGCACCGCCCGGGGATTACAACCCTCCACTG         7214

Qy   2221  CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGCCCACTGCCA              2280
           ||||||||||||  |||||||||||||| ||||||| ||||| ||||||||||| |||
Db   7215  CTAGGAGTTTCCCCGGCGGGACCCCGGGATTACGTCCCTCCAGTGGTACATGGGCCCCATTGCCA       7274

Qy   2281  CCTACTAAGACCCCTCCTATTACCACCTCACGGAGAACAGCAGTGTTCTGACAGAA                2340
           ||||| ||| |||| ||||||||||| ||||||||||||| ||| ||||||||||
Db   7275  CCTACCAAGGCCCCTCCAATACCACCTCACGGGGAGACGGTTGTTCCGGTTATCGGCGG             7334

Qy   2341  TCCACCGGTGTCTTCTGCCCTGGCCGGAGCTTGCCACAAAGGCTTTCGGTTAGCTTCCGAACCG        2400
           ||||||||||| | ||| ||||||||||||| | |||||||||||| |||| | | |||||||
Db   7335  TCCACCGGTGTCTTGGCAGCTCGCTACACCCTCCGTGGGCCCCCGGGATCG                     7394

Qy   2401  TCGGCCGTGACAGCGGGCAGGGCACGGGCCCCTCCTGACAACCGACGCGGA                     2460
           ||||| |||||||| |||| ||||| ||||||||| |||||||||||| ||
Db   7395  TCGGCCCGTTGACAGCGGGGCACGGGCCCCTCCGGACAACCAGTGAC                         7454

Qy   2461  GCAGGATCTGACGTTGAGGTCGTATTCCTCCTCCCGGAGCGGGGGAC                         2520
           |||||||| |||||||||||| ||| |||||||| ||||||||||||
Db   7455  GCAGGATCCCGACGTTGAGTGTATTCCTCCATGCCCCGGGGGAGCGGGGGGAT                    7514

Qy   2521  CCCGATCTCAGCGACGGGTCTTGGTTCTACGGTGAGGAGGTCGTC                           2580
           |||||| ||||||||||| ||| || || ||||||||||||||||
Db   7515  CCCGAATCTCAGCGACGGGTCTTGGTTCTACGGTGAGGAGGTCGTC                          7574

Qy   2581  TGCTGCTCGATGTCCTACACAGCGGACGCCATGCGCTGTGCGGAGGAA                        2640
           |||||||||||||| |||||| ||||||||||| ||||||||||||
Db   7575  TGCTGCTCGATGTCATACACATGGACGCCCTGATCACGCCTGCTGCCGGAGGGA                  7634

Qy   2641  AGCAAGCTGCCCATCAAACGGTTGAGCAACTCTTTGCTGCGGTCAACAACATGGTCTAC            2700
           ||||||| |||||||||| ||||||||||||| ||| |||   ||| || || ||| ||
Db   7635  AGCAAGCTGCCCATCAAATCGGTTGAGCAACTCTTTTGCTCACCAACAACATGGTCTAT           7694

Qy   2701  GCTACCACATCCCGGCAGGCCCAGCGCGGCAAGCCAGCAGGTCACCTTTGACAGACTGCAA         2760
           |||||||||||||||||||||||||||| ||||||||||||||||| |||||||||||||
Db   7695  GCCACAACATCCCGGCAGGCCGAAGAGAAGTCACTTTTGACAGACTGCAA                     7754

Qy   2761  ATCCTGGACGGATCACTACCAGGAGGCTGTGCTCAAAGGAGAATGAAGGCCGGTCCACAGTT        2820
           ||||||||||||||| ||||||||||||||||| |||| || | |||||||| ||||||||
Db   7755  ATCCTGGACGGATCACTACCAGGAGGCTGTGCTCAAGGAGATGAAGGCGTCCACAGTT            7814

Qy   2821  AAGGCTAAGCTTCTATCAGTAGGAGGAAGGCCTGCAGCTGCAGCTGCACCATTCCGGCCAAA       2880
           |||||| ||||| ||| || ||||| || || ||||||| | ||| ||||||||||||||
Db   7815  AAGGCTAAGCTTCTATCCGTGAGGAAGCCTGCCAGCTGCACCCACATTCCGGCCAAA             7874

Qy   2881  TCTTAAATTTGGCTARGGGGGCAAAGGAGCCTATCCAGAGGGGCCATTAACCAC               2940
           |||||| |||||| | || || ||||||  ||||||||| | |||| |||||||
Db   7875  TCTAAATTTGGCTATAGGGGGCAAAAGGACGTTCGGGAACCTATCCGCAAGGCCGTCAACCAC       7934

Qy   2941  ATCCGCTCCGTGTGGGAGGACACTGAAACACCAATTGACACCACCATC                    3000
           |||||||||||||||||| ||| ||||||||||||||||||||||||
Db   7935  ATCCGCTCCGTGTGGGAAGACTTGCTGGAAGACACCAAATTGACACCACCATC               7994

Qy   3001  ATGGCACAAAAATGAGGGGGTTTTTCTGGGTCCAACCAGAGAGGAGGAGGCCGCAAGCCAGCTCGC  3060
           ||||||||||||||||| || ||||||||||||||| ||||||||||||||||||||||||||||
Db   7995  ATGGCAAAAAAATGAGAGATTTTTCTGCTCCAACCAGAGAGGAGGAGGCCGCAAGCCAGCTCGC     8054
```

```
Qy   3061  CTTATGTGGTTCCCAGACTTGGGGGGTCCGGGTCGTGTTGTGCCGAGAAAATGGCCCTCTTGCCGGTG  3120
           |||||| ||||||||||||||| |||| || ||||| ||||||| | |||||     | | |||
Db   8055  CTTATCCGTTTCCCAGACTTGGGGGTTCCGTGTTGTGTGCCGAGAAAATGGCCCCTCTTACGATGTG     8114

Qy   3121  GTCTCCACCCTCCCTCCAGGCGTTGATCGGGGGTTCCGTACGGATTTCAGTAGTTCCCTGGA        3180
           ||||||||| |||||||||| ||||||||||||||||| |||||||||||||| |  |||||
Db   8115  GTCTCCACCTTCCTCAGGGCCGTGATGGGCGTCCTCATACGGATTTCAGTACTCCTCTGGA         8174

Qy   3181  CAGCGGGGTCGGAGTTCCTGGTGGAAACGCGGGTGGAAATCAAAAGGAAGACCCCTATGGGCCTTTGCA  3240
           ||| ||||||||||||||||||| | || | |||||||||||||| ||||| |||| | || || |||
Db   8175  CAGCCGGGTTGAGTTCCTGGTGGGATGCGCTGGATGCCTGGAAATCAAAGGAAGTGTCCTTATGGGCTTCGA  8234

Qy   3241  TATGACACCCCCTGTTTTGACTCAACAGTCACTGACGATGACACTCGGTGTGTAGAGGAGTCA        3300
           |||||||||||||| ||||| ||| ||||| ||| || | | ||||||||||| |||||||||
Db   8235  TATGACACCCCCGTGTTTTGACTCAACGGTTCAGCTGGATGCCTGGAAAATCAAAGAAGTGTCCTTATGGGCTTCGA  8294

Qy   3301  ATTTATCAAATGTGGGGGGGGTCCCCTGACTCGGCCCCCGGAAGCCGATCCGAGCTCACGAG        3360
           |||||| |||||||||| |||| ||||||||||||||| |||||||| | |||||||||||
Db   8295  ATCTACAATGTTGTGACTTGGCCCCCGGAGCCGATCCTCAACGGTGACTCCCGTCACGAGGTCGGTTATCGGCGG  8354

Qy   3361  CGGGTTTATATCGGGGGGGTCCCCTGACTAATTCAAAAGGGGCAAGACGTCCTTCACGAGGAA       3420
           |||||||| | ||||||||| ||||||||| |||||||||||||||||||||||| ||||
Db   8355  CGGGCTTTATGTCGGGGGGGTGCCCCTTGACTAATTCCAAGGGGCAAAACTGCGGGTTATCGGCGG     8414

Qy   3421  TGCCGCGGCGACGGGCCTGTCCTGCTACTAGCTGCGTGCGGTTAATACCCTCACATGTTACTTGAAG  3480
           ||||||||| ||||||| || |||||| || ||| |||  |||||||||| ||||| ||||||
Db   8415  TGCCGCGGCGGTGGCTGGCTGCCGCTACTAGCTGCGTGCAACACCCTCACTTCACATGTTACTTGAA   8474

Qy   3481  GCCTCTGCAGCCGCTGTCGAGCTGCAAAGCTCAAGGAGGATGTCGCAGATGCTCGGGAGAC         3540
           |||||||||||| ||||||||||||||| || ||| |||||||||||||||||| |||||
Db   8475  GCCTCTGCGCCTGTCGAGCTGCAAAGCTCCAGGACTGTCGCAGATGCTCGTGTGCTACGAGAC        8534

Qy   3541  GGCCTGTCGTGTTATTCTGTGAGAGCTGTGCCAGGACCCAGGAGGTCGGGGAC                 3600
           |||| ||||| || ||||| |||| |||| ||||||| ||||||||||||| |
Db   8535  GGACTCGTCGTTATTCTGTGGAAAGGCGGGTGCCAGGACCCAAGAGGATGGCTACGAGTC           8594

Qy   3601  TTCACCGGAGGGGCTATAGACTAGGTACTATTCTGCCCCCCGCCACCAGAATAC                3660
           |||||||||| | |  ||||| |||||||| ||| |||||||||| ||||||||
Db   8595  TTCACGGAGGCGTATGACTAGGTACTTACTCTGCCCCCCTGGGGGACGCGCTAACCGGGAATAC      8654

Qy   3661  GACCTGGAGTTGATAACATCATGCTCCTCTCCAAATGTGTCTCGCGGCCACGGATGCATCTGGC      3720
           ||||| |||||||| || || ||||| || ||||| ||| | ||||||| || ||||  |||
Db   8655  GACTTGGAGTTGATAACATCAAGCTCCTCCAAACGTGTCCCACGGATGCATCTGGC               8714

Qy   3721  AAAAGGGGTATACTACCTCACCCGTGACC  3749
           ||||||| ||||||| |||||| ||||||
Db   8715  AAAAGGGGTGTTACTATCTCACCCGTGACC  8743
```

RESULT 8
AB049090
LOCUS   AB049090   9573 bp   RNA   linear   VRL 22-AUG-2002
DEFINITION   Hepatitis C virus gene for polyprotein, complete cds,
             isolate:HCV11140.
ACCESSION   AB049090
VERSION   AB049090.1 GI:11559446
KEYWORDS   .
SOURCE   Hepatitis C virus
  ORGANISM   Hepatitis C virus
             Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
             Hepacivirus.
REFERENCE   1
  AUTHORS   Takahashi,K., Iwata,K., Matsumoto,M., Matsumoto,H., Nakao,K.,
             Hatahara,T., Ohta,Y., Kanai,K., Maruo,H., Baba,K., Hijikata,M. and
             Mishiro,S.
  TITLE   Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
             with hepatocellular carcinoma: the 'progression score' revisited
  JOURNAL   Hepatol. Res. 20 (2), 161-171 (2001)
  PUBMED   11348851
REFERENCE   2 (bases 1 to 9573)
  AUTHORS   Mishiro,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-SEP-2000) Shunji Mishiro, Toshiba General Hospital,

Department of Medical Sciences; 6-3-22 Higashi Oh-i, Shinagawa-ku,
Tokyo 140-8522, Japan (E-mail.shunji.mishiro@po.toshiba.co.jp.
Tel:81-3-3764-8981, Fax:81-3-3764-8992)

FEATURES   Location/Qualifiers

source   1..9573
    /organism="Hepatitis C virus"
    /mol_type="genomic RNA"
    /isolate="HCVT140"
    /specific_host="human"
    /db_xref="taxon:11103"
    /tissue_type="serum"
    /note="HCV-genotype: 1b"

5'UTR   <1..310
    /note="partial"

CDS   311..9343
    /codon_start=1
    /product="polyprotein"
    /protein_id="BAB18803.1"
    /db_xref="GI:11559447"
    /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRGPRL
    GVRATRKTSERSCQPRGRRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRG
    SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
    GVNYATGNLPGCPSIFLLALLSCLTIPASAVEVRNVSGVYHVTNDCSNSSIVYEAAD
    MIMHTPGCVPCVREDNVSRCWVALTPTLAARNSSIPTTTIRRHVDLLVGAAALCSAMY
    VGDLCGSVFLVSQLFTFSPRRYETVQTCNCSIYPGHVSGHRMAWDMMMNWSPTTALVV
    SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGNTHVTGGQA
    GQHAIRFTSLFSSGASQKIQLINTNGSWHINRTALNCNDSLHTGFIAALFYAHKFNAS
    GCPERMASCRPIDKFDQGWGPISHVVNISDQRPYCWHYAPRPCGIVPASQVGVYC
    FTPSPVVVGTTDRFGVPTYTWGENETDVLLLNNTRPPQGNWFGCTWMNGTGFTKTCGG
    PPCNIGGAGNTLTCPTDCFRKHPEATVTKCGSGPWLTPRCIVDYPYRLWHYPCTVNF
    TIFKVRMYVGGVEHRLNAACNWTRGERCNLEDRDRSELSPLLLSTTEWQIMPCSFTTL
    PALSTGLIHLHQNIVDVQYLYGVGSAVVSIVIKWEYVVLLFLLLADARVCACLWMMLL
    IAQAEAALENLVVLNAASVAGAHGTLSFLVFCAAWYIKGKLVPGAAYAFYGWPLLL
    LLLALPRAYAMDREMAASCGGTVFVGLVLLTLSPHYVFLARLIWWLQYFITRAEAC
    LQVWIPPLNVRGGRDAIILLTCAVHPELVFEITKILLAILGPLMVLQAGLTRVPYFVR
    AQGLIRACMLVRKVAGGHIVQWMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLVVAVE
    PVIFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSFGEQGWRLLAPIT
    AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSK
    TLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
    DSRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMETT
    MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
    GFGAYMSKAHGVDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIICDECHS
    TDSTSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
    AIPLEAIKGGRHLIFCHSKKKCDELAAKLSGLGVNAVAYYRGLDVSVIPTSGDVVVVA
    TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPFFIETTTVPQDAVSRSQRRGRTGRG
    RGTYRFVTPGERPSGMLDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQ
    DHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
    LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLA
    ALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYQEFDEMEECASHLPYIEQGMQLAE
    QFKQKALGLLQTATKQAERAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGN
    PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSI
    GLGKVLVDILAGYGAGVAGALVAFKIMSGETPSTEDLVNLLPAILSPGALVVGVVCAA
    ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
    RLHQWINEDCSTPCSGSWLRDVWDWICTVLVDFKTWLQSKVLPRLPGVPFLSCQRGYK
    GVWRGDGIMQTTCPCGAQIAGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPTPSP
    APNYSRALWRVAAEEYVEITRVGDFHYVTGMTTDNIKCPCQVPAPEFFTELDGVRLHR
    YAPACKPLLRDEVTFQVGLNQYVVGSQLPCEPEPDVTVLITSMLTDPSHITAETAKRRL
    ARGSPPSLASSASAQLSAPSLKATCTTRQDSPDADLIEANLLWRQEMGGNITRVESEN
    KVILLDSFEPLRAEEDEREVSVAAEILRKSRKFPPALPVWARDYNPPLIESWKDPDY
    VPPVVHGCPLPPTKAPPIPPPRKRTVVLTESTVSSALAELATKTFGSSGSSAVDSGT
    ATAPPDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLLSGGSWSTVSEEASEDVVCCSMS
    YTWTGALITPCAAEESKLPINALSNSLLRHHNWVYATTSRSASQRQKKVTFDRLQVLD
    DHYRDVLKEMKAKASTVAKAKLSVEEACKLTPPHSARSKFGYGAKDVRNLSSRAVNHI
    RSVWKDLLEDTQTPIDTTVMAKNEVFCVQPEKGGRKPARLIVTPDLESVTESDIRV
    EESIYQCCDLAPEARQAIRSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
    TCYLKATAACRAAKLQDCTMLVCGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPG
    DPPQPEYDLELITSCSSNVSVAHDASGKRVYLTRDPTTPLARAAWETARHTPVNSWL
    GNIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGATYSIEPLDLPQIIQRLH
    GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYL
    FNWAVRTKLKLTPIPAASQLDLSGWFIAGYSGGDIYHSLSRARPRWFMWCLLLLSVGV
    GIYLLPNR"

3'UTR   9344..9396

3'UTR   9397..9439
    /note="upstream of poly-U stretch"

3'UTR   9440..9573
    /note="poly(U) stretch"
    /note="poly(U) length is indistinct
    /note="3' X tail"

ORIGIN

Query Match   87.9%;   Score 3297.8;   DB 14;   Length 9573;
Best Local Similarity   92.5%;   Pred. No. 0;
Matches 3467;   Conservative   0;   Mismatches 282;   Indels   0;   Gaps   0;

```
Qy   3121 GTCTCCACCCCTCCCTCAGCTGTGTCGATGGGCTCCTCGTACGGGATTCCAGTATTCTCCTGGA 3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8102 GTCTCCACCCTTCCTCAGCCCGTGATGGGCCGCTCCATACGGGATTCCAGTATTCTTCCTGGA 8161

Qy   3181 CAGGGCGGGGTCGGAGTTCCTGGTGAACGCCTGGAAATCAAAGAAGACCCCTATGGGGCTTTGCA 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8162 CAGGGCGGGGTCGGAGTTCCTCCTGGAATGTCAAAGAAGACATCGGGGTCGGCA 8221

Qy   3241 TATGACACCCGCTGTTTGACTCAACAGTCACATCGTGTAGAGGAGTCA 3300
          ||||||||||||||||||||||||||||||||||||||||||||
Db   8222 TATGACACCCGCTGTTTGACTTGGCCCCGAAATCAAAGAGATCCGGGCCGCA 8281

Qy   3301 ATTTATCAAATGTTGTGACTTGGGCCCCCGAGACCACCGTCAGGCTGAGAGG 3360
          ||||||||||||||||||||||||||||||||||||||||||||
Db   8282 ATATACCAATGTTGGCCTCCCGAGACGGCCCATAAGGTTCGTCACACAGAG 8341

Qy   3361 CGGGCTTTATATCGGGGGTCCCTGACTAATTCAAAAGGGGCAGAACTGCCG 3420
          |||||||||||||||||||||||||||||||||||||||||||
Db   8342 CGGGCTTTATGTCGGGGTCCCCTGACTAATTCAAAAGGGCCATGGGCTATGCCCGG 8401

Qy   3421 TGCCGCGCCAGCGCGTGTGACCAACAGTGCTGCCTGCCTGTGCCTTTGAAG 3480
          ||||||||||||||||||||||||||||||||||||||||||
Db   8402 TGCCCGCCAAGCGGTGTGCTTGCCTAGCTGCCTTTGAAG 8461

Qy   3481 GCCTTCTGACAGCCTGTCGCAGCTGCAAAGCTTCCAGGACTGCAGATGGAC 3540
          |||||||||||||||||||||||||||||||||||||||||||
Db   8462 GCCACTGGGACTGGGATAACATCATGTCCTCCAACGTGTCGGCGTCGGAC 8521

Qy   3541 GGGCTTGTCGTTATCTGTGAGACGGCGGAACCCAGGAGCCTACGAGTC 3600
          ||||||||||||||||||||||||||||||||||||||||||
Db   8522 GACCTTGTCGTTATCTGTGAAAGCGCCGGAACCCAGGAGCCTACGAGTC 8581

Qy   3601 TTCACGGAGGCTATGACTAGGTACTCGGACCCGGGAGGGACCTTCCCGGG 3660
          ||||||||||||||||||||||||||||||||||||||||||
Db   8582 TTCACGGAGGCTATGACTAGGTACTCTGCCCCGGGGGACCCCCAGAATAC 8641

Qy   3661 GACCTGGGACTGGGATAACATCATGTATGCTCCCACTCTTTGGGCCAGGATGATCCTCATGACTCATTTCTTCTCCATCTTGCTGGC 3720
          |||||||||||||||||||||||||||||||||||
Db   8642 GACTTGGGACTGGGATAACATCATGTCCTCCAACGTGTCGGCGTCCGCCACGGTCCTGGC 8701

Qy   3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
          ||     ||   |||||||||||| |||
Db   8702 AAGCGGGTGTATTACCTCACCCGCGACCC 8730
```

RESULT 9

```
HCVPOLYP

LOCUS       HCVPOLYP      9379 bp    RNA     linear   VRL 15-JAN-1999
DEFINITION  Hepatitis C virus complete genome sequence.
ACCESSION   AJ000009
VERSION     AJ000009.1  GI:2764397
KEYWORDS    complete genome; core protein; E1 protein; E2 protein; NS2
            protein; NS3 protein; NS4a protein; NS4b protein; NS5a protein; NS5b
            protein; p7 protein; polyprotein.
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9379)
  AUTHORS   Trowbridge,R. and Gowans,E.J.
  TITLE     Molecular cloning of an Australian isolate of hepatitis C virus
  JOURNAL   Arch. Virol. 143 (3), 501-511 (1998)
  MEDLINE   98232263
   PUBMED   9572551
REFERENCE   2  (bases 1 to 9379)
  AUTHORS   Trowbridge,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1997) Trowbridge R.; Hepatitis Unit, Sir Albert
            Sakzewski Virus Research Centre, Royal Children's Hospital, Herston
            Road, Brisbane, Queensland Q 4029, AUSTRALIA
FEATURES             Location/Qualifiers
     source          1..9379
                     /organism="Hepatitis C virus"
                     /virion
                     /mol_type="genomic RNA"
                     /isolate="HCV-A"
                     /db_xref="taxon:11103"
     misc_RNA        1..9379
                     /note="genomic RNA for polyprotein precursor, complete
                     genome"
     5'UTR           1..324
     CDS             325..9357
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="CAA03854.1"
                     /db_xref="GI:2764398"
                     /db_xref="GOA:O93016"
                     /db_xref="UniProt/TrEMBL:O93016"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRALRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPSDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLGLLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAAD
                     MIMHVPGCVPCVRVDNSSRCWVALTPTLAARNASVPTTAIRRHVDLLVGAAATCSAMY
                     VGDLCGSVFLVAQLFTSPRHHTVQDCNCSIYPGHVSGHRMAWDMMMNWSPTAALV
                     SQLLRIPQAIVDMVAGAHWGVLAGLAYYSMVGNWAKVLIVMLLFAGVDGDTHTGGVA
                     GRDTLRFTGFFSLGPQQKIQLVNTNGSWHINRTALNCNDSLNTGWLAALFYSNFNAS
                     GCPERMASCHPIDEPAQGWRPITYAEHSSSDQRPYCWHYAPQPCGIVPASEVCGPVYC
                     FTPSPVVVGTTDRHGVPTYSWGENGTDVLLLNNTRPPQGNWFGCTWMNGTGFTKTCGG
                     PPCNIGGVGNNTLICPTDCFRKHPEATVTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TIFKVRMVVGGVEHRLSAACNWTRGERCDLEDRDRSELSPLLLSTTEWQVLPCSFTTL
                     PALSTGLIHLHQNIVDVQYLYGIGSVVSFAIKWEYVVLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAWYIKGRLVPGAAYAIYGVWPLL
                     LLLALPPRAYALDREMAASCGGAVFVGLVLLTLSPHYKEFLARLIWWLQYFITRAEA
                     LQVWIPPLNVRGGRDAIILLTCVVHPELIFDITKLLLAMLGPMVLQAVTKVPYFVR
                     AQGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLQDWAHGLRDLAVAVE
                     PVVFSDMETKIITWGADTAACGDIISGLPVSARGRILLGIISGLKQGWRILLAPIT
                     AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGASK
                     TLAGPKGPVTQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
                     DSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMETT
                     MRSPVFTDNSSPPAVPQTYQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGTDPNIRTGVRTITTGSPITYSTYGKFLADGGCSGGAYDIIICDECH
                     TDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNIGEIPFYGK
                     AIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYLNTPGLPVCQ
                     DHLEFWESVFTGLTHIDAHFLSQTKQAGENFPYLTAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSMWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGRPAIIPDRELVLQEFDEMEECASHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWQALEAFWAKHMWNFISGIQYLAGLSTLPGN
                     LAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSAAAAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAPKVMSGEMPESTHVVESDAAARVTQILSSLTIQLLK
                     ILRRHVGPGEAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTIQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGIPFYSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHR
                     YAPACKPLLRDEVTFQVGLNQYVVGSQLPCEPEPDVVVTSMLTDPSHITAEEFAKRRL
                     DRGSPPSLASSSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGGNIIRVESEN
                     KVVILDSFDPLRAEEDEREVSVAAEILRKSRRFPRAMPTWARDYNPPLLESWKDPDY
                     VPPVVHGCPLPPTKAPPIPPPRKKRTVVLTESTVSSALAELATKTFGSSESSSAVDSGT
                     ATAPPDQPDNDTGSDVESCSSMPPLEGEPGDPDLSDGSWSTVSEEASEDVVCCSMS
                     YTWTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKAKLLSVEEACKLTPHSARSKFGYGAKDVRNLSSKAVNHI
                     RSVWKDLLEDTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
                     VVSTLPQAVMGPSYGFQYSPGQRVEFLVNAWKSKKCPMGFAYDTRCFDSTVTESDIRV
                     EESIYQCCDLAPEARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
                     TCYLKASAACRAAKLQDCTMLVCGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPG
                     DPPQPEYDLELITSCSSNVSVAHDASGKRVYLTRDPTTPLARAAWETARHTPVNSWL
                     GNIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
                     GLSAFSLHSYSPGEINRVASCLRKLGVPPLRAWRHRARSVRAKLLSQGGRAANITGRYL
                     FNWAVKTKLKLTPIPAASQLDLSNWFVAGYSGGDIYHSLSRARPRWFMWCLLLSGV
                     GIYLLPNR"
     mat_peptide     325..897
                     /product="core protein"
     mat_peptide     898..1473
                     /product="E1 protein"
```

```
mat_peptide    1474..2562
               /product="E2 protein"
mat_peptide    2563..2751
               /product="p7 protein"
mat_peptide    2752..3402
               /product="NS2 protein"
mat_peptide    3403..5295
               /product="NS3 protein"
mat_peptide    5296..5457
               /product="NS4a protein"
mat_peptide    5458..6240
               /product="NS4b protein"
mat_peptide    6241..7581
               /product="NS5a protein"
mat_peptide    7582..9354
               /product="NS5b protein"
3'UTR          9358..9379
ORIGIN
```

Query Match   87.9%;   Score 3294.6;   DB 14;   Length 9379;
Best Local Similarity   92.4%;   Pred. No. 0;   Mismatches 284;   Indels 0;   Gaps 0;
Matches 3465;   Conservative 0;   Mismatches 284;   Indels 0;   Gaps 0;

```
Qy      1  TGGGAGGGGCTGTCTTCACAGGCCTCACCCACCGTGGATGCCCACTTCCTGTCCCAAACAAAG   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4996  TGGGAGAGGCGTCTTCACAGGCCTCACCCACATAGACGCGCACTTCTTGTTGTCCCAGACCAAA 5055

Qy     61  CAGGCAGGAGGACAACTTCCCTACCTGGTGGGGTACCAGGGCC                        120
Db   5056  CAAGGGGAGGAGAACTTCCCCTACCTGACGAAGGTACCAGGGCC                      5115

Qy    121  CAGGGCCCCACCTCATCCATCATGGGATCAAATGTGGAAGTGTTCTCATAAAAGCCTACT      180
Db   5116  CAGGCCTCACCTCACCTCATGTGGGATCAAATGTGTCTCATACGGGCTAAAAGCCTACG      5175

Qy    181  CTGCGCGGCGGGCCCAACACCCTTGCTGTTGTATAGGCTGGGGAGCCCTCCAAAAGACGA      240
Db   5176  CTGCACGGGGCGGGCCAACACCCCCTGCTGTTATAGGCTAGGACCGAGGTCGTTCCTT       5235

Qy    241  ACACACCCCATAACCAAAATTCATCATGGATGCATGTCAGCGGGCTGGGTCACG           300
Db   5236  ACACACCCCATAACCAAAATACATCATGGCATGTCGCGCTGACCTAGAGGTCGTCACG        5295

Qy    301  AGCACCCTGGGGTCGTGGTGGGGCGGGGCTCCTCGGCCTCGTTTGACAACA            360
Db   5296  AGACACCTGGGTGCTAGTGTGGGGGGAGTCCTCGCAGCCCCTGGCCTATTGCCTGACAACG    5355

Qy    361  GGCAGCGTGGCTGCTCATTGTGGGGTAGGATCATCTTTGTTCCGGGGACCCGGGCCAT     420
Db   5356  GGCAGCGGTGGTCATTGTGGGGCAGGATTATCTTGTTCCGGGAGCCCGGCTATCATTCCCGAC   5415

Qy    421  AGGGGAAGTGCCTCTACCAGAGTTCGATGAGATGAAGAAGAGTGCGGGTCGCCAC         480
Db   5416  AGGGGAAGTGCCTTTACCAGGAGTTCGATGAGATGGAGAGAGTGGCGCCTCACACCTTCCTTAC 5475

Qy    481  ATCGAGCAGGGAATGCAGCTTCGCGCCCAGCACTTCAAGCAAAAAGCCT              540
Db   5476  ATCGAACAGGGAATGCAGCTCGCCGGAACAATTCAAACACGAGGGCCTTCGGGTTGCTCGCAG   5535

Qy    541  ACACGCCACCAAAGCACGAGGACGCGCCGCCTGCTCCCGGTGGGTGGCGACCCCTT        600
Db   5536  ACAGCCACTAAGCAAGCGGAGGCTCGCTGTTCCCCGGGTTGGGGAAATCCGGCAAGCCCTT    5595

Qy    601  GAGACCTTCTGGGCCGAGAACACGTGGAACTTCATCACACGGGGGATAACAGTAC         660
Db   5596  GAGGCCTTTTTGGGCGGAGACATGTGGGAAGACATTCATCACGGGGGTGGAAGTTGGTCGCGTC 6195

Qy    661  TTGTCCACTCTGCCTGGGAATCCCGGCGATTGCATCACTGATGGGTTCACAGCCTTCTTC    720
Db   6736  AAACCCCTCCTACGGGATGAAGTCACATTCCAGGTGCTCACCAACTACGTGGTTGGG      6795

Qy    721  ACTAGCCCCCTCACCAACATTCTACCCCTCCTGCTTAACATCCTGGGGGGGGGATGGGTAGCC   780
Db   6796  TCACAACTCCCATGCGGAGCCCGAACCGGATGGTAGTGCTGGGGTGGTTCAGTTAGCGAC     6855
```

Right column:

```
Db   5716  ACCAGCGCCGTCCACCACCAACATACCCTCCTGTTTAACATCTTGGGGGGTGGGGGCT       5775
Db   6016  TGCGCAGCGATACTGGGGTCCGGCACGTGGGTCCCAGGGAGGCGGGCCTTCGGGGGATGAAC   6075

Qy   1021  TGCGCAGCGATACTGGGGTCCGGCACGTGGGTCCCAGGGAGGCGGGCCTTCGGGGGATGAAC   1080
Db   6016  TGCGCAGCGATACTGGGGTCCGGCACGTGGGTCCCAGGGAGGCGGGCCTTGGGGGGATGAAC  6075

Qy    961  GACCTGGTTAACTTACTCCCTGCCATCCTCTCTCCTCGGGGGTCGGGTCGTCGGGGTCGTG   1020
Db   5956  GACCTGGTCAACTTACTACCCTGCCATCCTCTCTCCTCGGGGGTCGGGTCGTCGGGGTCGTG   6015
```

[Sequence alignment continues]

```
Qy       1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920
              ||||||||||| |||||||||||||||||||| ||||||| |||||||||||||||||||
Db       6856 CCCTCCCACATTACAGCAGAGACGGCTAAGCGTAGGCTGGACAGGGGGTCTCCCCCCTCC 6915

Qy       1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980
              |||||||||||||||||||||||||||||||||||||||| ||||||||||| || ||||
Db       6916 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTTGAAGGCGACATGCACTACC 6975

Qy       1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040
              |   | |||| |||||||||||||||||| ||||||||||||||||||||||| ||||||
Db       6976 CGTCACGACTCCCCAGACGCTGACCTCATTGAGGCCAACCTCCTGTGGCGGCAGGAGATG 7035

Qy       2041 GGCGGGGACATTACCCGCGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
              |||||  |||| |||||||||||||| || |||||||||||||| |||||||||||||||
Db       7036 GGCGGAAACATCACCCGCGTGGAGTCTGAAAACAAGGTAGTAATTCTGGACTCTTTCGAC 7095

Qy       2101 CCGCTCCGAGCGGAGGAGGATGAGCGGGAAGTGTCCGTCCCGGCGGAGATCCTGCGGAAA 2160
              ||||| |||||||||||||||||| ||||||| |||||| |||||||||| |||||||||
Db       7096 CCGCTTCGAGCGGAGGAGGATGAGAGGGAAGTATCCGTCGCGGCGGAGATTCTGCGGAAA 7155

Qy       2161 TCCAAGAAATTCCCACCAGCGATGCCCGCATGGGCACGCCCGGATTACAACCCTCCGCTG 2220
              |||| || |||||| | ||||||||||  ||||||||| |||||||||||||| || |||
Db       7156 TCCAGGAGATTCCCCCGAGCGATGCCCATATGGGCACGGCCGGATTACAACCCCCCACTG 7215

Qy       2221 CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGTGCCCACTGCCA 2280
              || |||||||||||||  ||||||||||||||||| ||||| || ||||||||| | |||
Db       7216 CTAGAGTCCTGGAAGGATCCGGACTACGTCCCTCCGGTGGTGCACGGGTGCCCATTACCA 7275

Qy       2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340
              ||||| ||| ||||||| ||||||||||||||||  |||||||| ||||| |||||||| 
Db       7276 CCTACCAAGGCCCCTCCAATACCACCTCCACGGAAAAAGAGGACGGTTGTCCTGACAGAG 7335

Qy       2341 TCCACCGTGTCTTCTGCCCTGGCGGAGCTTGCCACAAAGGCTTTCGGTAGCTCCGAACCG 2400
              |||||||||||||||||| ||||||||||||| |||||| | ||||| ||||||||| ||
Db       7336 TCCACCGTGTCTTCTGCCTTGGCGGAGCTTGCTACAAAGACCTTCGGCAGCTCCGAATCG 7395

Qy       2401 TCGGCCGTCGACAGCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACGACGGCGGA 2460
              ||||||||||||||||||||||| ||||||||||||||||| ||| ||||| ||| ||  
Db       7396 TCGGCCGTCGACAGCGGCACGGCGACCGCCCCTCCTGACCAGCCCCCCGACAACGACGAC 7455

Qy       2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAC 2520
               ||||||| |||||||| ||||  ||||||||||||||||||||||||||||||||||| 
Db       7456 ACAGGATCCGACGTTGAATCGTGCTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAT 7515

Qy       2521 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGTGAGGAGGCCGGTGAGGACGTCGTC 2580
              ||||||||||||||||||||||||||||||||||| ||||||||  ||||||||||||||
Db       7516 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGCGAGGAGGCTAGTGAGGACGTCGTC 7575

Qy       2581 TGCTGCTCGATGTCCTACACATGGACAGGCGCTCTGATCACGCCATGCGCTGCGGAGGAA 2640
              ||||| |||||||||||||| ||||| |||||||||||||| |||||||| ||||| |||
Db       7576 TGCTGTTCGATGTCCTACACGTGGACGGGCGCTCTGATCACACCATGCGCCGCGGAAGAA 7635

Qy       2641 AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTAC 2700
              ||||||||||||||||| |||||||||||||||||||||||||| ||||||||||| |||
Db       7636 AGCAAGCTGCCCATCAATGCGTTGAGCAACTCTTTGCTGCGTCATCACAACATGGTGTAC 7695

Qy       2701 GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAA 2760
              || || || ||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       7696 GCCACAACCTCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACTTTTGACAGACTGCAA 7755

Qy       2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 2820
               |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db       7756 GTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 7815

Qy       2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAAA 2880
              |||||||| |||||||| ||||| ||||||||||||||||| |||||||||||||||| |
Db       7816 AAGGCTAAACTTCTATCCGTAGAAGAAGCCTGCAAGCTGACACCCCCACATTCGGCCAGA 7875

Qy       2881 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 2940
              |||||||||||||| ||||| |||||||||||||||||||||||||||||| ||||||||
Db       7876 TCTAAATTTGGCTACGGGGCGAAGGACGTCCGGAACCTATCCAGCAAGGCCGTTAACCAC 7935

Qy       2941 ATCCGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000
              ||||||||||||||| |||||||| ||||||||||||||||||||||||| || ||||||
Db       7936 ATCCGCTCCGTGTGGAAGGACTTGCTGGAAGACACTGAAACACCAATTGATACTACCATC 7995

Qy       3001 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGAGGAGGCCGCAAGCCAGCTCGC 3060
              |||||||| |||||||| ||||||||||||||||| | ||||||||||||||||||||||
Db       7996 ATGGCAAAGAATGAGGTCTTCTGCGTCCAACCAGAAAAAGGAGGCCGCAAGCCAGCTCGC 8055

Qy       3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGTGTGTGCGAGAAAATGGCCCTCTATGACGTG 3120
              |||||||||||||||||||||||||| || || |||||||| ||||| || |||||||||
Db       8056 CTTATCGTGTTCCCAGACTTGGGGGTGCGCGTATGCGAGAAGATGGCTCTTTATGACGTG 8115

Qy       3121 GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTACGGATTCCAGTATTCTCCTGGA 3180
              ||||||||||| |||||||| ||||||||| ||||||||||||| ||||| |||||||||
Db       8116 GTCTCCACCCTTCCTCAGGCCGTGATGGGCCCCTCGTACGGATTTCAGTACTCTCCTGGA 8175

Qy       3181 CAGCGGGTCGAGTTCCTGGTGAACGCCTGGAAATCAAAGAAGACCCCTATGGGCTTTGCA 3240
              |||||||||||||||||||| || ||||||||||||||||||   ||||||||||| |||
Db       8176 CAGCGGGTCGAGTTCCTGGTAAATGCCTGGAAATCAAAGAAGTGTCCTATGGGCTTCGCA 8235

Qy       3241 TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300
              |||||||||||||||||||||||||| |||||||||| |||||||||||| |||||||||
Db       8236 TATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAGTGACATCCGTGTTGAGGAGTCA 8295

Qy       3301 ATTTATCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 3360
              ||||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db       8296 ATTTACCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAAGTCGCTCACAGAG 8355

Qy       3361 CGGCTTTATATCGGGGGTCCCCTGACTAATTCAAAAGGGCAGAACTGCGGCTATCGCCGG 3420
              |||||||| ||||||||||||||||||||||||||||||||||||||||| |||||||| 
Db       8356 CGGCTTTACATCGGGGGTCCCCTGACTAATTCAAAAGGGCAGAACTGCGGTTATCGCCGA 8415

Qy       3421 TGCCGCGCGAGCGGCGTGCTGACGACTAGCTGCGGTAATACCCTCACATGTTACTTGAAG 3480
              |||||||| ||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       8416 TGCCGCGCAAGCGGCGTGCTGACGACTAGCTGCGGTAATACCCTTACATGTTACTTGAAG 8475

Qy       3481 GCCTCTGCAGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTCGTGTGCGGAGAC 3540
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8476 GCCTCTGCGGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTCGTGTGCGGAGAC 8535

Qy       3541 GGCCTTGTCGTTATCTGTGAGAGCGCGGGAACCCAGGAGGACGCGGCGAGCCTACGAGTC 3600
              | ||| |||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db       8536 GACCTCGTCGTTATCTGTGAAAGCGCGGGAACCCAAGAGGACGCGGCGAGCCTACGAGTC 8595

Qy       3601 TTCACGGAGGCTATGACTAGGTACTCTGCCCCCCCCGGGGACCCGCCCCAACCAGAATAC 3660
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       8596 TTCACGGAGGCCATGACTAGGTACTCTGCCCCCCCCGGGGACCCGCCCCAACCAGAATAC 8655

Qy       3661 GACCTGGAGTTGATAACATCATGCTCCTCCAATGTGTCGGTCGCGCACGATGCATCTGGC 3720
              ||||||||| ||||||||||||||||||| |||||||||||||||||||||||||| |||
Db       8656 GACCTGGAGCTGATAACATCATGCTCCTCGAATGTGTCGGTCGCGCACGATGCATCCGGC 8715

Qy       3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
              || || |||||||||||||||||||||||
Db       8716 AAGAGAGTATACTACCTCACCCGTGACCC 8744
```

```
RESULT 10
AF207753
LOCUS       AF207753                9374 bp    RNA     linear   VRL 27-APR-2000
DEFINITION  Hepatitis C virus strain MD12 complete genome.
ACCESSION   AF207753
VERSION     AF207753.1  GI:7650223
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9374)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Miyasaka,Y., Marumo,F. and
            Sato,C.
  TITLE     Characteristics of hepatitis C viral genome associated with disease
            progression
```

JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 9374)
  AUTHORS   Nagayama,K., Kurosaki,M., Enomoto,N., Miyabaka,Y., Izumi,N. and
            Sato,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku,
            Tokyo 113-8519, Japan
FEATURES   Location/Qualifiers
     source   1..9374
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="MD12"
                     /db_xref="taxon:11103"
                     /note="isolated from a patient with persistently-normal
                     ALT values (ASC-2)"
     CDS   330..9362
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF65943.1"
                     /db_xref="GI:7650224"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPGRRQPIPKVRQPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARVLAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTTPASAYEVRNVSGMYHVTNDCSNSSIVYEAAD
                     MIMHAPGCVPCVRENNSSRCWVALTPTLAARNASVPTTTIRRHVDLLVGAAAFCSAMY
                     VGDLCGSVFLISQLFTFSPRRHTVQDCNCSIYPGHVSGHRMAWDMMMNWSPTAALVV
                     SQLLRIPQAVVDVVVAGAHWGVLAGLAYYSMQGNWAKVLIVMLLFAGVDGHTRISGGTQ
                     GHTTSGLVSLFSSGSSQNIQLINTNGSWHINRTALNCNDSLNTGFIAALFYTKHFNAS
                     GCPARMASCRSIDTFAQGWGPITYAQPSISDQRPYCWHYAPRQCGIVPASQVCGPVYC
                     FTPSPVVVGTTDRRGVPTYSWGENETDVLLLNNTRPPRGNWFGCTWMNSTGFTKTCGG
                     PPCNIGGVGNNTLTCPTDCFRKHPEATVTKCGSGPWLTPRCLVDYPYRLWHYPCTVNF
                     SIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQVLPCSFTTL
                     PALSTGLIHLHQNVVDVQYLYGIGSAVVSFVVKWDYVVLLFLLLADARVCACLWMMLL
                     IAQAEAALENLVLINAASVAGAHGLFSLVFFCAAWYIKGRLVPGAAYALYGVWPLLL
                     LLLALPPRAYAMDREMAASCGGAVLVGLMILTLSPYYKVFLARLIWWLQYILTRAEAH
                     LQVWVPPLNVRGGRDAIIFLTCAHTPELIFDTKILLLAILGPLMVLQAGITRVPYFVR
                     AQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVDHLTPLRGWAHTGLRDLAVAVE
                     PVVFSDMTKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGWRLLAPII
                     AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQTFLATCVNGVCWTVFHGAGSK
                     TLAGPKGPITQMYTNVDQDLVGWQAPQGARSMTPCTCGSSDLYLVTRHADVIPVRRRG
                     DGRGSLLSPRPVSYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAVMSKAYGTDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFYGK
                     AIPIETIKGGRHLIFCHSKKKCDELAAKLSALGINAVAYYRGLDVSVIPTSGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIDTTVPQDAVSRTQRRGRTGRG
                     RGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQ
                     DHLEFWEGVFTGLTRDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LTRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVVTSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGRPAIIPDREVLYREFDEMEECASQLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWRALEAFWGKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASVTSPLTTQSTLLFNIIGGWVAAQLAPPSASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVGVVCAA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSNLITQLLR
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLSDFKTWLQSKLLPRLPGVPFLSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTSP
                     APNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHR
                     YAPACKPLLREEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAATAKRRL
                     ARGSPPSLASSSASQLSAPSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVVILDSFDPLRAEEBEREISVPAEILRKPKRFFPALPIWARPDYNPPLLESWKDPDY
                     VPPVVHGCPLPPTKPPPRKRTVVLTESTVSSALAELATKTFGSSGSSAADSGT
                     ATAPPDQTSDDGDKGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEAGEDVVCCSMS
                     YTWTGALITPCAAEEEKLPINALSNSLLRHHNMVYATTSRSACQROKKVTFDRLQVLD
                     DHYRDLVKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAVDLTGPLGVRVCEKMALYD
                     RSVWKDLLEDETPINTTVMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
                     VVSTLPQAVMGSSYGFQYSPGQRVEFLVNTWKSKKCMGFSYDTRCRCEDSTVTENDIRV
                     EESIYQCCDLAPEARQVIRSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
                     TCYLKASAACRAAKLQDCTMLVNGDDLVICESAGTQEDAASLRVFTEAMTRYSAPPG
                     DPPQPEYDLELITSCSSNVSVAHDAGSKRVYLLTRDPTTPLARAAWETAFHTPVNSWL
                     GNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLH
                     GLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGRAAICGKYL
                     FNWAVRTKLKLTPIPAASRLDLSGWFVAGYSGGDIYHSLSRARPRWFMLCLLLSVGV"

ORIGIN

Query Match "GIYLLPNR"
Best Local Similarity 92.4%;
Matches 3464; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
Score 3293; DB 14; Length 9374;
Pred. No. 0;

Qy   1 TGGGGAGGGGGTCGTTCACCAGGCCTCACCCACCGTGGATGGGCCCACTTCCTGTGTCCCAAACAAAG   60
           ||||||||||| || ||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db   5001 TGGGGAGGGGGGTCTTTTACAGGCCTCACCCGCCGTGGATGAGGCCCACTTCCTGTGTCCCAGGGCC   5060

Qy   61 CAGGCAGGAGACAACTTCCCCTACTGGTCGGCGTGGCTACCAAGGCTACTGTGTGCGCTAGGGCC   120
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   5061 CAGGGCAGGAGGACAACTTCCCCCTACATGGGTAGGTAGCCATACCAAGGCCACGGGTATGGGCC   5120

Qy   121 CAGGCCCCACTTCCTCCATCCATGGGGATCAAAATGTGGGAAGTGTTCTCATACGGGCTAAAGCCTACT   180
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5121 CAGGGCTTCCACTTCCATCGTGGGCGACCAAAATGTGGGCTGCCTTCCACACGGGCTACGCCTACT   5180

Qy   181 CTGCGCGGCGGGGCCAACACCCTTGCCTTGTATAGGCCTGGGGAGCGCGGTCCAAAAACGAGGGTCACCCTC   240
           ||||||||||||||||||||||||||||| |||||||||| |||| ||||||||||||||||||||||||||
Db   5181 TTGCACGGGGCCAAACGCCCCCTGCTATATAGGCCTGGGGAGCGCGGTCCAAAAATGAGGGTCACCCTC   5240

Qy   241 ACACACCCCCATAACCAAAATTCATCATGGCCATGCATGTCAGCCGACCTGGAGGTCGTACG   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5241 ACACACCCCCATAACCAAAATACATCATGACACATGCCATGCATGCTCGGTGGGATAGGT   5300

Qy   301 AGCACCTGGGTGCTGCTGGTCGCGGGGTCCGCGGGTCTGCCTGCCTGCTGTTGCTTGACAACA   360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5301 AGCACCTGGGTGCTGGTAGGGCGGAGTTCTTCGCAGACTCGGGCTGCCTGCCTGGACAACA   5360

Qy   361 GGCAGCCTGGGTCATTGTGGTGTAGGATCATCTTGTCCGGGGAGAGCCTATTGTTCCCGGAC   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5361 GGCAGCCTGGTGTCATTGTGGGTAGGGATCATCTTGTCCGGGGAGAGCCCCATTATTCCCGGAT   5420

Qy   421 AGGGAAGTCCTTCTACCAGGAGTTCGATGAGATGGAAGAGAGTGCGGTGCGCACCTCCCTTAC   480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5421 AGGGAAGTCCTTCTACCAGGGAGTTCGATGAAAGTGGGGGATACAGTACTTCCTTAC   5480

Qy   481 ATCGAGGACAGGGAAATCGAGCTCGCGCCCAGAGCCAGTTCAGGCAAAAAAGCCCTCGGGTTGTCAG   540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5481 ATCGAACAGGGAAATGCAGCTCGCCGGAGCAAATCAAACAGAAGAGGGCGTTGCTGCAA   5540

Qy   541 ACAGCCACCAAGCACGGAGGAGCCGCTGCTCCGGTGGGAGGAGCCGAGCCCTT   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5541 ACAGCTACCAAGCAAGCGGAGGCTGCTCCGGTCTGTTAAACATCTTAGCAGGC   5600

Qy   601 GAGACTTCGCTCCCCGAAACACATGTCGAACTTCATCAGCGGCGATACAGTACTTAGCAGGC   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5601 GAGGGCCTTCTGGGGAAGCACATGTGGAATTCATTAGCGGGGATACAGTACTTAGGACAGGC   5660

Qy   661 TTGTCCACTCTGCCTCTGGGGAATCCCGCCGGATTGCATCGACGATGGGCCTTCACAGCCTCTGTC   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5661 TTGTCCACTCTGCCTCTGGGAACCCCGCAATAGGCCATCACACTGATGGCCCTGTC   5720

Qy   721 ACTAGCCCGTCACCAACATCTACCCTCCTGCTTAACATCCTGGGATGGGTAGCC   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5721 ACCAGCCCGGCTCACCACACCCAAAGTAACCTCCTGTTTAACATCTTGGGGATGGGT   5780

Qy   781 GCCCAACTCGCTCCCCCAGTGCTTCAGCTTTCGTTAGGCGCCATTGCTGGTGCG   840
           |||| ||||||| || |||||||||||||||||| |||||||||||||||||||
Db   5781 GCCCAACTCGCTCCCCCAGCGCTCGGCCTTCGGCCTTGTGGGGACTAGGGCGG   5840

Qy   841 GCTGTTGGGCAGCATAGGGCCTTGGGAAGGGTCGTTGTGGGACAATCTTCGGCGGCGCTATGGAGCA   900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5841 GCTGTAGGCAGCCATAGGCCCTTGGGGCCTCGGGCGGGTATTGGCCCGGTATGCGGCA   5900

Qy   901 GGAGTGGCAGGCGGCGCCTGCTTGCCTTAAGGTCATAGAGCGGGGAAATGCCCTCACCGAG   960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5901 GGAGTGGCTGGGCGGCCACTCGGTGGCCCTTTAAGGTCATGGAGCGGCCGAGATGCCCTCACTGAG   5960

Qy   961 GACCTGGTTAACTTACTCCCTGCCATCCTCTCCGGGTCGCCTCGGTCGGTGCGGGGGTCGTG   1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5961 GGAGTGGGCTGGGCCACTCGTGCGGCTCTTTAAGGTCATGAGCGGCGGGAGATGCCCTCCACTGAG   1020

Db   5961 GACCTGGTTAATTTACTCCCTGCATCCTCTCCTGGGCGGCCCTTGGTCGGGGGTCGTG 6020
Qy   1021 TGCCGCAGCCGATACTGCCGTCGGCCACGTGGGGTCCAGGGAGGGCTGTGCAGTGGATGGAAC 1080
Db   6021 TGGCGCAGCAATACTGCGGTCGGGACGTGGGGCTGTGCAGTGGGAATGAAC 6080
Qy   1081 CGGCTGATAGGGTTTCGCGCTCGCGGGGTAACCATGTTTTGCCACAGAG 1140
Db   6081 CGGGCTGATAGGCGTTGCGCCTGCGGGGGTAACCACGGTCTCCCCCAGGCACTGTTGCCTGAG 6140
Qy   1141 AGCCAACGCCGCCAGGAGCGCAGTGTCACTCAGAGATCCTTCTCCGAACTTACTATCACCCAACTGTTG 1200
Db   6141 AGCCAACGCCGCCAGACGTTACTCAGAGATCCTTCCCAACCTTACCATCACTCAGCTGCTG 6200
Qy   1201 AAGAGGCTCCACCAGTGGGATTAACCGGAGGACTGCTTCCAGGGGCCCTGCTCGGGCTCGGTGGTCTA 1260
Db   6201 AGGAGGCTTCACCAGTGGGATTAATGAGGACTGCTTCCACCGGGCCATGGCTCCGGGCTCGTGGTCTA 6260
Qy   1261 AGGGATGTTTGGGGACTGGGATATGCACAGTTTTTGGCTTCAAGGACCTGGCTCCAGTCC 1320
Db   6261 AGGGGATGTTTGGGGACTGGGATATGCACGGTCGTTGTCGACTTCAACCTGGCTCCAGTCC 6320
Qy   1321 AAGCTTCCTGCCGCGGCGGATTACGGGGAGTCCCCTTTTCTCAAGCGTTGGGGTACAAGGGG 1380
Db   6321 GTCTTGGCGCGGGGAGGGGCATCATGGCGAGTCATCGCAAAACCACTGTTCCTCTCCGCACAGATCACCGGA 6380
Qy   1381 CATCTCAAAAACGGTTCCATGAGGATGGTCCCCTTTCTCTCAGTCAGTTACATGTTGGCAT 1440
Db   6381 GTCTGGGCGGGGAGGGGCATCATGCCGGAGTCGTCCATGTTACATGTTGGCAT 6440
Qy   1441 CATGTTCAAGAAGCGGTTCCATGAGGATTGTTGGTCTGACTTCAACCTGGCTCCAGTCC 1500
Db   6441 CATGTTCAAGAAGCGGTTCCATGAGGATTGTTGGGCCCTTAAAAACCTGGCAGCAACACGTGGCAT 6500
Qy   1501 GGAACATTCCCCATCAAACGGATACACCACCACGCCCCTGCAGCGCCCCCAGCGCCCAAAC 1560
Db   6501 GGAACATTCCCCATCAAACGGATACACCACCACGCCCCCTGCACACCCCCAGCGCCCAAAC 6560
Qy   1561 TATTCCAGGGCGCTGTTGGGCGGTGGCTGCTGCTGCTGAGGAGGTTACGCGGGGTGGGG 1620
Db   6561 TATTCTAGGGCGCTTGGGCGGTGGCTGCTGCTGCTGAGGAGGTTACGCGGGGTGGGG 6620
Qy   1621 GATTTCCACTACGTGACGAGCACACTGACACAGTAAAAATGCCCGGTGCCCAGGTTCCA 1680
Db   6621 GATTTCCACTACGTGACGGGCATGACCACACTGACAACAGTAAATGCCCATGCCAGGTCCCG 6680
Qy   1681 GCCCCCCGGAATTCTTCACGAGAAGTGGAATGGGGTGCGCGGCTGCACAGTTAGGTCCCGGCGTGC 1740
Db   6681 GCCCCCGAATTCTTCACTGAGGTGATGGAGTGCGCGGCTGCACAGTTAGCGTCCGGCGTGC 6740
Qy   1741 AAACCTCTTCCTTACTAGGGGAGGAGGAGTCACATTCCAGTTCAGGCTCAACCAATACCTGGTTGGG 1800
Db   6741 AAACCTCTTCCTTACGGGGGAGGGGAGTCACATTCCAGTTCCTGGTTGGG 6800
Qy   1801 TCGCAGCTCCCATGGGCAGCCGGATAGTAGCAGTGTCTCCAGTCGTCTCACTTCCATGCTACCACGAC 1860
Db   6801 TCACAGCTCCCATGCAGCCGGAACCGGACGTGCCCAGTGCTCACTTCCATGCTTACGGAC 6860
Qy   1861 CCCTCCCCACATCACAGCAGAGACAGCGGCTAGGCGCCAGGCGGGGGTTCTCCCCCTCC 1920
Db   6861 CCCTCCCACATTACAGCAGAGACAGCGGCTAAGGCGGGGTTCTCCCAGTTCTCCCCCCTCC 6920
Qy   1921 TTGGGCCAGCTCTTCAGCTAGCCAGTTGTTCTCGCGCCCTTCCTGAAGGCGACATGCACTACC 1980
Db   6921 TTGGGCCAGCTCTTCAGCTAGCCAGTTGTTTCTGCGCGCCATGCACTACC 6980
Qy   1981 CAAAAATGACTTCCCAGACGCTGACTCATCGAGGCCTAACCCCTCCTTGAAGGCGCATGAGATG 2040
Db   6981 CACCAATGACTCCCCGGACGCTGACTCATCGAGGCCAACCCCTCTTACGGAGATG 7040
Qy   2041 GGCGGGGGACATTACCCGCGGTTGGGAGTCAGAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
Db   7041 GGCGGGAACATCACCCGCGTGGGAGTCAGAGAATATCTGGGACTCTTTTGGAT 7100

Qy   2101 CCGCTTCCGGACCGGAGGGAGGAGGAGGAGGATGAGCGGGGAAGTGTCGGTCCGGCCGAGATCCTGCGGGAAA 2160
Db   7101 CCGGCTTCCGGCCGGGGACGGAGGATGAGAGGGAGAAAATATCCGTTCCGGGGGATCCTGCGGGAAA 7160
Qy   2161 TCCAAGAAATTCCCACCCACCAGGATGCCCCGCATCCGGAGGACGCCCCGGATTACAACCCTCCGCTG 2220
Db   7161 CCCAGGAAGTTCCCCCCCCACGCGTTGCCCCATATGGGCGGGATTACAACCCTCACTG 7220
Qy   2221 CTGGAGTCCTGGAAGGCCCCGGACTTACGGTCCCTCCAGTGGGTACATGGGGTGCCCACTGCCA 2280
Db   7221 CTAGAGTCCTTGGAAGGACCCCGGACTACGTCCCTCCAGTGGGTTGCACGGGTGCCCATTGCCA 7280
Qy   2281 CCTACTAAGACCCCTCCTTATACCACCTCCACGGAGAAGGAGGAGGAGGACAGTTGTTCTGACAGAA 2340
Db   7281 CCTACCAAGGGCCTTCCATTACCACCTCACCGGACACCTGTCCTAACAGAG 7340
Qy   2341 TCCACCGGTGTCTTCTGCCCCTGGGCGGGGAGGCTTGGCCACAAAGGCTTTGGGGTAGCTCCGAACCG 7400
Db   7341 TCCACCGGTGTCTTCTGCCCCTGGGCGGGGGGCTTGACTTATAAGGACCTTAGGCGACCTCCG 7400
Qy   2401 TCGGCCGTCGACCAGCGGCCACGGGGCCACACCGGGCCCCTCCTGACAAACCCTCGACGAGCGGGA 2460
Db   7401 TCGGCCGACCGGACAGCAGGGGCCACGTGGCCACGCCCCTCCTGGACTCAGCAGCGACGGAC 7460
Qy   2461 GCAGGATCTGACGTTCGAGTGGTATTCCTCACGGGGCCCCCCCCCCTTGGGGGAGCGGGGGGGAC 7520
Db   7461 AAAGGATCCGACGTTGAGTGTACTCCTCCAATGCCCCCCCCCTTACTAGGACCTTAGGCTTAGTCG 7520
Qy   2521 CCGGATCTCAGCGGACGGGTTCTTGGTTCTACCGGTGAGTGAGGGAGAAGGAGCGGGGGGGAC 7580
Db   7521 CCGGATCTCAGCGATGGGTGTCTTGGTTCTACCGTGACGGAGAAGCCTTGCTCGTC 7580
Qy   2581 TGCCTGCTTCGGATGTGTCCTACACATGGACAGGGGGCTCTGATCAGCGACATGCGTGCGGAGGAA 7640
Db   7581 TGCTGCTTCAATGTCCTTACCACGGAGTTTAGGCCATGCCGCTGCCGGAGGAA 7640
Qy   2641 AGCAAGCTGCCCATCAACCGGGTTTGAGCCAACTTCTTTGCTGCGGCTCACCACACCATGGTCTAC 7700
Db   7641 AGCAAGCTGCCCATCAAACGGCCATGCAGAGGCGCAGAAGAAAGTCACCTTTGACAGACTGCAA 7700
Qy   2701 GCTACCACCACTCCCGCGAGGCCAAGGCCCAGAGGAAGGTCACCTTTGACAGACTGCAA 7760
Db   7701 GCTCACAACATCCCGCAGCCTGAGCCATTCTTTGACAGACTGCAA 7760
Qy   2761 ATCCTGGACGCAATCACTACCACCAGGACGTGCAAAGGAAGCAGTTGAAGGCGGAGGCGAAAGGGTG 7820
Db   7761 GTCCTGGACGACCACCTACCCGGGACGTGCTCAAAGGAGATGAAGGGTGGTCCACAGTG 7820
Qy   2821 AAGGCTAAGCTTCTATCAGTAGGATGAGGAAGCCTTGCACGCTGCCCCCCACCACTTTGGCGCCAAA 7880
Db   7821 AAGGCTAAACTTCTATCTGTAGAGGAAGCTTGACGCCCCACATTTGGCGCCAAA 7880
Qy   2881 TCTAAATTTGGCTATGGGGGGCAAAAGGAGCGTCCGGGAACCTTATCCAGCAAGGCGCCATTAACCAC 7940
Db   7881 TCCAAAATTTGGCTTACGGGGGCAAAAGACGTTCGGGACGCCTTATCCAGCAAGGCGGTTAACCAC 7940
Qy   2941 ATCCGCTCCGTGTGGGAGGGACTTGGTTGGGAGGACACTGAAACACCAATTGACAACGGTC 8000
Db   7941 ATCCGCTCCGTGTGGGGAGGGACTTGCTTGGGGAGGACACTGAAACCAATTAACACCACCATC 8000
Qy   3001 ATGGCAAAAAAATGAGGTTTTCTGCGGTCCAACCAGAGAGAGCGGCCGCAGCCGAGCTTCGC 8060
Db   8001 ATGGCAAAAAATGAGGTTTTTCTGCGGTCCAACCAGAGAAAGGAGCGGCCCGAAAACCCGGCTCGC 8060
Qy   3061 CTTATCCGTTCCCAGACTTGGGGGTCCGGTGTGTGGTGGGGAGAAAAATTGACACCACCACCATG 8120
Db   8061 CTTATCGTTATTCCCACAGACCTGGAGTTCGGTGTGTTTGGGGGGAAGAGATGGAGTG 8120
Qy   3121 GTCTCCACCCTCCCCAGGGTTGATGGGGCTTCCTCGGTACGGCAAATTCCAGTATTCTCCTGGA 3180
Db   8121 GTTTCCACCCTTCCTCAGGGCGCCGTGATAACGGGATTCCACAGTACTCTCCTGGG 8180

RESULT 11
AR488101

```
LOCUS       AR488101                8642 bp    DNA     linear   PAT 15-MAY-2004
DEFINITION  Sequence 2 from patent US 6706874.
ACCESSION   AR488101
VERSION     AR488101.1  GI:47253849
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 8642)
  AUTHORS   Kukolj,G. and Pause,A.
  TITLE     Self-replicating RNA molecule from hepatitis C virus
  JOURNAL   Patent: US 6706874-A 2 16-MAR-2004;
FEATURES             Location/Qualifiers
     source          1..8642
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

Query Match   87.8%; Score 3292.2; DB 6; Length 8642;
Best Local Similarity   92.3%; Pred. No. 0;
Matches 3462; Conservative 2; Mismatches 285; Indels 0; Gaps 0;

```
Qy        1 TGGGAGGGGCTGTCTCACAGGCCTCACCACCAGTGGATGGCCCCACTTCCGTGTCCCAAACAAAG     60
            |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     4046 TGGGAGGAGGCGTCTTTACAGGCCCTCACCACCACATAGAGCGCCCATTTCTTGTGCCCAGACTAAG  4105

Qy       61 CAGGCAGGAGGAGACAACTTCCCCTCACCTGGTGGGCGGTACCAGGCTACGTGTGTTCGCGGCTGGCT    120
            ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4106 CAGGCAGGAGAGACAACTTCCCCTACCTGGTAGCATACCAGGCTACGGTGTGCGGCCAGGGGCT      4165
```

(The remainder of the page consists of a continuous Qy/Db nucleotide sequence alignment continuing through positions 121–5260, presented in paired "Qy" query and "Db" database rows with vertical bar match indicators.)

```
RESULT 12
LOCUS       AX472292        8642 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 2 from Patent WO2005015.
ACCESSION   AX472292
VERSION     AX472292.1  GI:22207307
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Kukolj,G. and Pause,A.
  TITLE     Self-replicating rna molecule from hepatitis C virus
  JOURNAL   Patent: WO 02052015-A 2 04-JUL-2002
            BOEHRINGER INGELHEIM (CANADA) LTD. (CA)
FEATURES             Location/Qualifiers
     source          1..8642
                     /organism="Hepatitis C virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11103"
     CDS             1802..8407
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD43568.1"
                     /db_xref="GI:22207308"
                     /translation="MDREMAASCGGAVFVGLILLSPHYKLFLARLIWWLQYFITRA
                     EAHLQWIPPLNVRGGRDAVILLTCAHPELIFTITKILLAILGPLMVLQAGTKVPY
                     FVRAHLIRACMLVRKVAGGHVVQMALMKLAALTGTVVYDHLTPLRDWAHAGLRDLAV
                     AVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQWRLLA
                     PITAYSQGTRGLLGCIITSLTGRDQNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGA
                     GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR
                     RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESM
                     ETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVA
                     ATLGFGAYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDE
                     CHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPF
                     YGKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVI
                     VVATDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRT
                     GRGRMGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLP
                     VCQDHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQM
                     WKCLIRLKPTLHGPTPLLYRGAVQNEVTTHPITKYIMACMSADLEVVTSTWVLVGG
                     VLAALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQ
                     LAEQFKQKAIGLLQTATKQAERAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLSTL
                     PGNPAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAV
                     GSIGLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVV
                     CAAILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQ
                     LLKRLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVFFSCQR
                     GYKGVWRGDGIMQTTCPCGAQITGHVRNGSMRIVGPRTCSNTWHGTFPINAYTTGPCT
                     PSPAPNYSRALMRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVR
                     LHRYAPACKPLLREEVTFLVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAK
                     RRLARGSPPSLASSSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGGNITRVE
                     SENKVVILDSFEPLQAEEDEREVSVPAEILRRSRKFPRAMPIWARPDYNPPLLESWKD
                     PDYVPPVVHGCPLPAKAPPIPPPRRKRTVVLSESTVSSALAELATKTFGSSESSAVD
                     SGTATAPSDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSSEASEDVVCC
                     SMSYTWTGALITPCAAEEKLPINALSNSLLRHHNLVYATTSRSASKFGYGAKDVRCHA
                     VLDDHYRDVLKEMKAKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRCHA
                     NHIRSVWKDLLEDTEPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMA
                     LYDVVSTLPQAVMGSSYGFQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTEND
                     IRVEESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCG
                     NTLTCYLKAAAACRAAKLQDCTMLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSA
                     PPGDPKPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVN
                     SWLGNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQ
                     RLHGLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCG
                     KYLFNWAVRTKLKLTPIPAASQLDLSSWFVAGYSGGDIYHSLSRARPRWFMWCLLLLS
                     VGVGTYLLPNR"
ORIGIN
     variation       4446
                     /note="r = a or g"
     variation       6268
                     /note="r = a or g"
```

```
Qy   3421  TGCCGCCGCAGCGCGGCGGTGCTGCTGACGACTAGCTGCGGTAATACCCTCACACATGTTACTTACTTGAAG 3480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7466  TGCCGCCGCAGCGCGGCGGTGTACTGACCGAAC                                         7525

Qy   3481  GCCCTCGCAGCCTGCGAGCTGCAAAGCTCAGAGACTGCACGAATACCCTCGTCTGCGTGCGGAGAC 3540
Db   7526  GCCGCTGCGCGCCTGTCGAGCTGCGAAGCTCCAGGACTGCACGAATGCTCCGGAGAC          7585

Qy   3541  GGCCTTGTCGTTATCTGTGAGAGCGGCGGAACCCAGGAGGACGCGGCGCCTACGGAGTC 3600
Db   7586  GACCTTGTCGTTATCTGTGAAAGCGGGGACCCAAGGAGGACGCGAGCCTACGGCC       7645

Qy   3601  TTCACGGAGGAGCGCTATGACTAGGTACTCTCGCCCGGGGGGAACCCAGGAATAC 3660
Db   7646  TTCACGGAGGAGCGCTATGACTAGATACTCTCGCCCCCCCCCAAAACCAGAATAC  7705

Qy   3661  GACCTGGAGTTGATAACATCATGCTCCTCCCAAATGTGTCGGTCGGCCGCACGGATGCATCTGGCC 3720
Db   7706  GACCTGGAGTTGATAACATCATGCTCCTCCCAAATGTGTCAGTCGGCGCGCACGATGCATCTGGC    7765

Qy   3721  AAAAGGGTATACTACCTCACCGTGACCC 3749
Db   7766  AAAAGGGTGTACTATCTCACCGTGACCC 7794
```

```
Query Match    87.8%;  Score 3292.2;  DB 6;  Length 8642;
Best Local Similarity  92.3%;
Matches 3462;  Conservative  2;  Mismatches 285;  Indels  0;  Gaps  0;

Qy     1  TGGGAGGGGGTGCTTCACAGGGCCTCACCCACTGGATGGCCCCACTTCCTGTCCCAAACAAAG 60
Db  4046  TGGGAGAGCGTGCTCTTTACAGGCCCTCACCCACCACAGACGCCCATTCTTGTCTGGCTAGGCCAGTA 4105

Qy    61  CAGGCAGGAGGAGACAACTTCCCCTACCTGGTGCGCGTACCAGCCGTACTGTGTGGCTAGGGCG 120
Db  4106  CAGGCAGGAGGAGACAACTTCCCCTACCTGGTAGCATACCAGGCTACGGTGTGGGCGCCAGGGCT 4165

Qy   121  CAGGCCCCACCTCCATCACATGGGATCAAAATGTTGGAAGTGTGTCTGCATCACGGCTAAAGCCTACT 180
Db  4166  CAGGCTCCACCTCCATCGTGGGACAAAATGTGGAAGTGTCTCATACGGGCTAAAGCCTACG 4225

Qy   181  CTGCGCGGGGGCGCAACACCCTTGCTGTATAGGCCGTCCAAAAACGAGGTCACCCTC 240
Db  4226  CTGCGACGGGGCGCAACGCCCCCTGCTGTATAGGCTGGGAGCCGTCCAAAAACGAGGTCGTCACG 4285

Qy   241  ACACACCCCATAACCAAATTCATCATGGCATGCATGTCAGCCGACCTGGAGTCGTCACG 300
Db  4286  ACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTCACG 4345

Qy   301  AGCACCTGGGTGCTCGTGGGCGGGGTCCTTGTGCAGTCTGGCTCGGTATTCGCTTGACAACA 360
Db  4346  AGCACCTGGGTGCTGGTGGGCGGAGTCCTTGCGGCTGCCTGGCGTTGGCCTGACAACA 4405

Qy   361  GGCAGCGTGGTCATTGTGGGTAGGATCAATCACTTGTCCGGGGCGGCCGGGCGGGTTTCCGGAC 420
Db  4406  GGCACCGTGGTCATTGTGGGCAGGATCATCCTTGTCCGGGGAAGCGCCGGCCATCCCGAC 4465

Qy   421  AGGGAAGTCCTCTACCAGACTTCGATGAGATGGAAGAGTCCGCGGTCGCGCGCCTTCCCTTAC 480
Db  4466  AGGGAAGTCCTTTACCGGGAGTTCGATGAGATGGAAGAGTGCGCCTCACACCTCCCTTAC 4525

Qy   481  ATCGAGCAGGGAAATCCAGTCTCGCGCGGAGCAGTTCAAGCAAAAAGCGGCTCCGGGTTGCTGCAG 540
Db  4526  ATCGAACAGGGAAATGCAGCTCGCCGAACAATTCAAAACAGTGGGCAATCGGGTTGCTGCAA 4585

Qy   541  ACAGCCACCAAGCAAGCGGAGGCCCCTGCTCCCGTGGTGCAGGACCATCTGGAGTTCTGGGAGTCC 600
Db  4586  ACAGCCACCAAGCAAGCGGAGGCTGCTGCCCGAACAATTCAAGTGGGCAGAGCCCTC 4645

Qy   601  GAGACCTTCTGGGGCCGAAACACATGTGGACTTCATCAGCGGGGTGGTGATAGTACAGTTAGCAGGC 660
Db  4646  GAAGCCTTCTGGGGCGAAGCATATGTGGAATTCATCAGCGGGATCACAATATTCAGCAGGC 4705

Qy   661  TTGTCCACTCTGCCTGGGAATCCCGCGGATGCGACACACTGATTGGCGTTCAGCAGCCTCGTC 720
```

```

```
               ||||||||||||||||| || |||||||| |||||||||||||| |||||||||||| ||
Db   4706 TTGTCCACTCTGCCTGGCAACCCCGCGATAGCATCACTGATGGCATTCACAGCCTCTATC 4765

Qy   721 ACTAGCCCGCTCACCACCCAATCTACCCTCCTGCTTAACATCCTGGGGGGATGGGTAGCC 780
               || |||||||||||||||||| |||||||||| |||||||||||||||||||||| |||
Db   4766 ACCAGCCCGCTCACCACCCAACATACCCTCCTGTTTAACATCCTGGGGGGATGGGTGGCC 4825

Qy   781 GCCCAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTAGGCGCCGGCATTGCTGGTGCG 840
               |||||||| ||||| ||||| |||||||| |||||||||||||||||||| ||||| |||
Db   4826 GCCCAACTTGCTCCTCCCAGCGCTGCTTCTGCTTTCGTAGGCGCCGGCATCGCTGGAGCG 4885

Qy   841 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGACATCTTGGCGGGCTATGGAGCA 900
               |||||||||||||||||||||||||||||||||||||| || ||||| || |||||||||
Db   4886 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGATATTTTGGCAGGTTATGGAGCA 4945

Qy   901 GGAGTGGCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGGCGAAATGCCCTCCACCGAG 960
               || ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   4946 GGGGTGGCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACCGAG 5005

Qy   961 GACCTGGTTAACTTACTCCCTGCCATCCTCTCTCCTGGTGCCCTGGTCGTCGGGGTCGTG 1020
               |||||||||||| |||||||||| |||||||| ||||| ||||| |||||||||||||||
Db   5006 GACCTGGTTAACCTACTCCCTGCTATCCTCTCCCCTGGCGCCCTAGTCGTCGGGGTCGTG 5065

Qy   1021 TGCGCAGCGATACTGCGTCGGCACGTGGGTCCAGGGGAGGGGGCTGTGCAGTGGATGAAC 1080
               ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   5066 TGCGCAGCGATACTGCGTCGGCACGTGGGCCCAGGGGAGGGGGCTGTGCAGTGGATGAAC 5125

Qy   1081 CGGCTGATAGCGTTCGCCTCGCGGGGTAACCATGTTTCCCCCACGCACTATGTGCCAGAG 1140
               ||||||||||||||||| |||||||||||||| || |||||||||||||||||||| |||
Db   5126 CGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCTGAG 5185

Qy   1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTCTCCGACCTTACTATCACCCAACTGTTG 1200
               |||||||| |||||||||||||||||||||||||| ||||| ||||| || ||| ||
Db   5186 AGCGACGCTGCAGCACGTGTCACTCAGATCCTCTCTAGTCTTACCATCACTCAGCTGCTG 5245

Qy   1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCACGCCCTGCTCCGGCTCGTGGCTA 1260
               |||||||| ||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db   5246 AAGAGGCTTCACCAGTGGATCAACGAGGACTGCTCCACGCCATGCTCCGGCTCGTGGCTA 5305

Qy   1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
               || ||||||||||| ||||||||||| || ||| |||| |||||||||||||||||||||
Db   5306 AGAGATGTTTGGGATTGGATATGCACGGTGTTGACTGATTTCAAGACCTGGCTCCAGTCC 5365

Qy   1321 AAGCTCCTGCCGCGATTACCGGGAGTCCCCTTTTTCTCATGCCAACGTGGGTACAAGGGG 1380
               ||||||||||||||||| |||||||||||||| |||||||| |||||||||||||||||
Db   5366 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 5425

Qy   1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440
               ||||||||||| |||||||||||||| ||||||||| |||||||||||||||||||||||
Db   5426 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 5485

Qy   1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
               ||||| |||||| |||||||||||||||| ||||||| |||||||||||||| |||||||
Db   5486 CATGTGAAAAACTGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5545

Qy   1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560
               |||||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||
Db   5546 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 5605

Qy   1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5606 TATTCTAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 5665

Qy   1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACGTAAAATGCCCGTGCCAGGTTCCA 1680
               |||||||||||||||||| |||||||||||||||||||||| |||||||| ||||||||
Db   5666 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5725

Qy   1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTGC 1740
               |||||||||||||||||||||||||||||||||||| |||||||||||||||| ||||||
Db   5726 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGCTCCAGCGTGC 5785

Qy   1741 AAACCTCTCCTACGGGAGGAGGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTTGGG 1800
               ||||| ||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db   5786 AAACCCCTCCTACGGGAGGAGGTCACATTCCTGGTCGGGCTCAATCAATACCTGGTTGGG 5845

Qy   1801 TCGCAGCTCCCATGCGAGCCCGAACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 1860
               || |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   5846 TCACAGCTCCCATGCGAGCCCGAACCGGACGTAGCAGTGCTCACTTCCATGCTCACCGAC 5905

Qy   1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920
               ||||||||||| || || |||||||||||||| |||||||||||||| ||||||||||||
Db   5906 CCCTCCCACATTACGGCGGAGACGGCTAAGCGTAGGCTGGCCAGGGGATCTCCCCCCTCC 5965

Qy   1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980
               ||||||||||| |||||||||||| ||||||||||||||| |||||| |||| || ||||
Db   5966 TTGGCCAGCTCATCAGCTAGCCAGCTGTCTGCGCCTTCCTTGAAGGCAACATGCACTACC 6025

Qy   1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040
               | |||||| ||| |||||||||||||||||||||||||||||||||||||| ||||||
Db   6026 CGTCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATG 6085

Qy   2041 GGCGGGGACATTACCCGCGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
               |||||| |||| ||||||||||||||||| || ||||||||||| |||||||||||||
Db   6086 GGCGGGAACATCACCCGCGTGGAGTCAGAAAATAAGGTAGTAATTTTGGACTCTTTCGAG 6145

Qy   2101 CCGCTCCGAGCGGAGGAGGATGAGCGGGAAGTGTCCGTCCCGGCGGAGATCCTGCGGAAA 2160
               ||||||| |||||||||||||||| ||||||| ||||| |||||||||||||||||||
Db   6146 CCGCTCCAAGCGGAGGAGGATGAGAGGGAAGTATCCGTTCCGGCGGAGATCCTGCGGAGG 6205

Qy   2161 TCCAAGAAATTCCCACCAGCGATGCCCGCATGGGCACGCCCGGATTACAACCCTCCGCTG 2220
               |||| ||||||||| | |||||||||| ||||||||||||||||||||||||||| |||
Db   6206 TCCAGGAAATTCCCTCGAGCGATGCCCATATGGGCACGCCCGGATTACAACCCTCCACTG 6265

Qy   2221 CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGTGCCCACTGCCA 2280
                |:||||||||||||| ||||||||||||||||||||||||||| ||||| ||| ||||
Db   6266 TTRGAGTCCTGGAAGGACCCGGACTACGTCCCTCCAGTGGTACACGGGTGTCCATTGCCG 6325

Qy   2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340
               ||| | ||| ||||||| |||||||||||||||||||||||||| ||||| ||| |||||
Db   6326 CCTGCCAAGGCCCCTCCGATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGTCAGAA 6385

Qy   2341 TCCACCGTGTCTTCTGCCCTGGCGGAGCTTGCCACAAAGGCTTTCGGTAGCTCCGAACCG 2400
               || ||||||||||||||| |||||||||| ||||||||| | ||||| ||||||||| ||
Db   6386 TCTACCGTGTCTTCTGCCTTGGCGGAGCTCGCCACAAAGACCTTCGGCAGCTCCGAATCG 6445

Qy   2401 TCGGCCGTCGACAGCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACGACGGCGGA 2460
               |||||||||||||||||||||||||| ||| |||||||||| ||||||||||||||||
Db   6446 TCGGCCGTCGACAGCGGCACGGCAACGGCCTCTCCTGACCAGCCCTCCGACGACGGCGAC 6505

Qy   2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAC 2520
               || ||||| |||||||||||||| |||||||||||||||||||||||||||||||||||
Db   6506 GCGGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAT 6565

Qy   2521 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGTGAGGAGGCCGGTGAGGACGTCGTC 2580
               |||||||||||||||||||||||||||||||| || |||||||| ||||||||||||||
Db   6566 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTAAGCGAGGAGGCTAGTGAGGACGTCGTC 6625

Qy   2581 TGCTGCTCGATGTCCTACACATGGACAGGCGCTCTGATCACGCCATGCGCTGCGGAGGAA 2640
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   6626 TGCTGCTCGATGTCCTACACATGGACAGGCGCCCTGATCACGCCATGCGCTGCGGAGGAA 6685

Qy   2641 AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTAC 2700
               | ||||||||||||||| || |||||||||||||||| |||||||||||| |||||||
Db   6686 ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCCGTCACCACAACTTGGTCTAT 6745

Qy   2701 GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAA 2760
               ||||| ||||| ||||||||||||| |||||||||||||||||||||||||||||||||
Db   6746 GCTACAACATCTCGCAGCGCAAGCCTGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAG 6805

Qy   2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 2820
                |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db   6806 GTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 6865

Qy   2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAAA 2880
               |||||||| |||||||| || ||||||||||| ||||||||||||||||||||||||| |
Db   6866 AAGGCTAAACTTCTATCCGTGGAGGAAGCCTGTAAGCTGACGCCCCCACATTCGGCCAGA 6925
```

RESULT 13
LOCUS   AR406043   7989 bp   DNA   linear   PAT 18-DEC-2003
DEFINITION Sequence 10 from patent US 6630343.
ACCESSION   AR406043
VERSION   AR406043.1 GI:40155170
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 7989)
  AUTHORS   Bartenschlager,R.
  TITLE   Hepatitis C virus culture system
  JOURNAL   Patent: US 6630343-A 10 07-OCT-2003;
FEATURES   Location/Qualifiers
     source   1..7989
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Query Match 87.8%; Score 3291.4; DB 6; Length 7989;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;
```

```
Qy     1 TGGGAGGGGGGCGTTCTTCACGAGGCCTCACCCACCGGTGGATGCCCACTTCCTGTCCCAAACAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3397 TGGGAGACGGGTCGTTACGAAGCCTCACCCACACATAGACGCCACCATTTCTTGTCCAGACTAAG 3456

Qy    61 CAGGCGAGGAGAGACAACTTCCCCTACCTGGTGCGGTACCAGGCTACTGTGTGGCCTAGGCGC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3457 CAGGCAGGGAGGACAAACTTCCCCTACCTGGTAGGCATACCAGGCTACGGTGTGGCCCAGGGCT 3516

Qy   121 CAGGCCCCACCTCCCATCATGTGGGAACCAAAATGTGGCAGCTGCAAAAGCCTACT 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3517 CAGGCTCCACCTCCATCGTGTGGGACCAAAATGTGGCTCTCATACGGCCTACG 3576

Qy   181 CTGGCGCGGGGCCAACACCCTTGCCTTGTATAGGGTGGGGAGCGCCTC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3577 CTGCGCAGGGGCCAAACGCCCCTGTCATTGTATAGGCTGGGGAGCCGTTCAAAAACGAGGTTACTACC 3636

Qy   241 ACACACCCCATAACCAAATTCATCATGGCATGCATGTCAGCCGACCTCACAGAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3637 ACACACCCCATAACCAAATACATCATGGGTGGCCACCTCACAGAG 3696

Qy   301 AGCACCTGGGTGCTCGTGTGGCGGGGCGGGGGTCCTTGCAGCTCTGCCTGCGTATTGCTTGACAACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3697 AGCACTGGGTTGCTGGTGAGGCGGGGAGTCCTAGCGCTTCTGGCGGCTCGGTGTTGCCTGACAACA 3756

Qy   361 GGCAGCCGTGGTCATTGTGGGGTTAGGATCATCTTGGTCCGGGAGTTGTTCCCGAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3757 GGCAGCGGGTTGGTCATTGTGGGCAGGATTCATCATCTTGGCTCGGGAGGGGCCATCATTCCCGAC 3816

Qy   421 AGGGGAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCGCGGTTGCTCCCTTAC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3817 AGGGGAAGTCCTTTACCGGGAGTTCGATGAGGAGATGGAGGAAGAGTGCGGCGCCCTCACCTCCTTAC 3876

Qy   481 ATCGAGCGAGGAAATGGAGCTCGCGCCGAGCAGTTCAAGGCAAAAAGCGCCTCGGGGTTGCTGCAG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3877 ATCGAACAGGGAATGCAGCTCGCCGAACAATTCAAACAGAGGGCAATCGGGTTGCTGCAA 3936

Qy   541 ACAGCCACCAAGGCAGCAGGGGAGGCCGGTTGCTCCGGTGGGCTCGAAGTTGGCAGGCCCTT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3937 ACAGCCACCAAGGCAAGCAAGCGGAGGCTGCGCTCCCCGTGGTGGAATCCAAGTGGCGGACCCTC 3996

Qy   601 GAGACCTTCTGGGCGAAACACATGTCGGAACTTCATCAGGTGGGATACAGTACTTAGCAGGC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3997 GAAGCCTTCTGGGCGGAGCATATGTGGAATTTCATCAGCGGGATACAAATAATTTAGCAGGC 4056

Qy   661 TTGTCCACTCTGCCTTGGGGAATCCCGGATTGCATCACTGATGGCAGTTCACAGCCTCTGTC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4057 TTGTCCACTCTGCCTCGCCAACCCCGGATAGCACTCACTGATGGCATTCACAGCCTCTATC 4116

Qy   721 ACTAGCCCGGCTCACCACCCCAAATCTACCCTCCTTGCTTAAACATCACCGGGGGGGAGGGAGGGGG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4117 ACCAGCGCCGGCTCACCACCAACATACCCCTCCCTGTTTAAACATCCCTGGGGGGGATGGGGTAGCC 4176

Qy   781 GCCCAAACTCGCTCCCCCAGTGCTTGCTTCCGTTTCGTTAGGCGCACGGCCCGGCATTGTCGTGCG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4177 GCCCAACTTGCTCCTCCCAGCCGCCTGCTTTCGTAGGCGCGGCCGGCATCGCTGGACGG 4236

Qy   841 GCTGTTGGCAGCATAGGCCTTGGGAAGGGTGCTTGTGGACATCTTGGCGGCGCCTATGGAGCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4237 GCTGTTGGCAGCATAGGCCTTGGGAAGGGTGCTTGTGCGGATATTTTTGGCAGGTTATGGAGCA 4296
```

```
Qy  2881 TCTAAATTTGGGCTATTGGGGCGGGGCAAAGGACGGTCCGGGAACCTTATCCCAGCAGCAAGGGCCATTAACCAC 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6926 TCTAAATTTGGGCTATTGGGGCGGGGCAAAGGACGGTCCGGGAACCTTATCCCAGCAGCAAGGGCCGTTAACCAC 6985

Qy  2941 ATCCGGCTCCGTGTGGGGAGGACTTGTTGTTGGGAGACACTGAAAACAACAATTGACACCACCATC 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6986 ATCCGGCTCCGTGTGGGAAAGGACTTGCTGGGAAGGACACTGAGGACACCAAGGACACCACCATC 7045

Qy  3001 ATGGCAAAAAATGAGGTTTTCTTGCCTCCACCAGAGGACGAGGCCGCAAGCCACACTCGC 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7046 ATGGCAAAAAATGAGGTTTTCTTGGCGTCCAACCAGAGAAGGGGCCGCAAGCCAGCCGC 7105

Qy  3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGTTGTCGTGTCGGGGAGAAAATGGCCCCTTATGACGTG 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7106 CTTATCGTTATTCCCGADATTTGGGGGTTCGTGTGGTCGGGAGAGAAGGGCCGCAGCCCAGCGTTCGATGTG 7165

Qy  3121 GTCTCCACCCTCCCTCAGGCGTGTGATGAGGGGGTCCTCGTGACGTATTCTCCTGGA 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7166 GTCTCCACCCTCCCTCAGGCGGCGTGATGGGGCTCTTCATACGGCCTTCATCTCCTGGA 7225

Qy  3181 CAGGCGGGGTCGAGTTCCTGGTGAAAGCCCTTGAAAATCAAAAGAGAGACCCCCTTATGGGGCTTTTGCA 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7226 CAGGCGGGGTCGAGTTCCTGGTGAATGCCTGGAAAAGCGCAAGAAAATGCCCTATGGGGGCTACG 7285

Qy  3241 TATGCACACCCGGCTGCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTGTGAGAGGAGTCA 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7286 TATGACACCCGGCTGCTGTTTGACTCAACAGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 7345

Qy  3301 ATTTATCAATGTTGTGACTTGGCCCCCCGAGCGCCATAGGGTGCTGCTCACAGAG 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7346 ATCTACCAAATGTTGTGACTTGGGCCCCCGGAAGCTCATAGGGTCGCTGGCTCACAGAG 7405

Qy  3361 CGGGCTTTATATCGGGGGGTTCCCCTGACTAATTCAAAAAGGGCCAGGAACTGGGCTATCGGCCCG 3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7406 CGGGCTTTACATCGGGGGGCCCCCTGACTAATTCTAAAAGGGCAGAACTGCGCGGGTCTATCGGCGG 7465

Qy  3421 TGCCGCGGCGGCGGGGGCGGTGCTAGCGTGACGACTAGTCTGCGGTAATACCCTCACATCCCTCACATCTTGAAG 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7466 TGCCGGCGGGGGGGGTGTCTGACGACCACCGCCTGTAATACCCTCACACATGTTAGTTACTTGAAG 7525

Qy  3481 GCCTTCTGCAGCCTGTCGAGCTGCAAAGCTTCCAGGACTGCACGATCTCGTTGTGCCGAGAC 3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7526 GCCCCTGGGCGGCCTGTCGAGCTGCCGGAAGCTCCAGGAGCTTCGGTTATGCGGGAGAC 7585

Qy  3541 GGGCCTTGTCGTTATCTGTGAGAGCGGCGGCGGGAAACCCAGAGGAGGACGAGGCGGGGCGGGGGGCCTTACGAGTC 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7586 GACCTTGTCGTTATCTGTGAAAGCGGGGACCAAGAGACCCTACGGGGCC 7645

Qy  3601 TTCACGGAGGGCTATGACTAGGGTACTCGCCCCCGGGGGGGGGGGGACCCGCGCCCCCCCCAACCAGAATAC 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7646 TTCACGGAGGGCTATGGAGTACTCTGCGCCCCGGGGGGACCGCTCGGTGAAGTGGCGGAGCCCTC 7705

Qy  3661 GACCTGGAGTTGATAAACATCATGCTCCTCCAAATGTGTCGGTCGCCGCGCCCAAAACCAGGC 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7706 GACTTGGAGTTGATAACATCATGCTCCTCCAAATGTGTCACTGGCCGCCAGATTGCATCTGGC 7765

Qy  3721 AAAAGGGGTATACTACCCTCACCCGTGACCC 3749
           |||||||||||||||||||||||||||||||
Db  7766 AAAAGGGTGTACTATCTCACCCGTACCC 7794
```

```
RESULT 14
LOCUS       AX036255              7989 bp    DNA     linear   PAT 16-NOV-2000
DEFINITION  Sequence 4 from Patent EP1043399.
ACCESSION   AX036255
VERSION     AX036255.1  GI:11225871
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Bartenschlager,R.D.
  TITLE     Hepatitis C virus cell culture system
  JOURNAL   Patent: EP 1043399-A 4 11-OCT-2000;
            BARTENSCHLAGER RALF DR (DE)
FEATURES             Location/Qualifiers
     source          1..7989
                     /organism="Hepatitis C virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11103"
ORIGIN

Query Match             87.8%;  Score 3291.4;  DB 6;  Length 7989;
Best Local Similarity   92.4%;  Pred. No. 0;
Matches 3463;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;
```

```
Qy       1 TGGGGAGGGCGGTTGGCTCTTCACAGGGCCTCACCCGTGGATGCCCACTTCCTGTGTG 60
           ||||  || |||||||||||||||||||||||||||||||||||| |||||||||||||
Db    3397 TGGGGAGGAGCGTTTTTACAGGCCTCACCACACATAGAGACGCCCATTTCTTGTGCTAAAG 3456

Qy      61 CAGGGAGGAGACAACTTCCCCTACCTTGGTGCGTACCAGGCTACTGTGTGCGGTAGGGCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3457 CAGGGAGGAGACAACTTCCCCTACCTTGGTAGCATACCAGGCTACTGTGTGCGGTAGGGCT 3516

Qy     121 CAGGGCCCCACCTCCATCATCATGGGGATGTAAAATTGGGAAGTGTTGTTCTATATGCCTACT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3517 CAGGGCTCCACCTTCACTCATCGTGGGGACCAAAATGTGGGAAGTGTTCTCATAGGCTTACG 3576

Qy     181 CTGCGCGGGGGGGCCAACACCCTTGCTCGTGTATAGGCGTGGGGGAGCGCCCAAAACGAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3577 CTGCACGGGCCAAACGCCCTGCTATTGTATAGGGGTGGGGGAGCCGTTCAAAAACGAGGCT 3636

Qy     241 ACACACCCCATAACCAAAATTCATCATGGGCATGCATGTCAGCCGACCTTGAGGAGGTCA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3637 ACACACCCCATAACCAAAATACATCATGGGCATGGTCCTGACCTGAGCTTGGAGGTCGTCACG 3696

Qy     301 AGCACCTGGGTTGCTGGTGTAGGCGGGGAGTCCTACCAGCTTCTGGGCGCGGTATTGCCTCACG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3697 AGCACCTGGGTGCTGGTAGGCGGGGAGTCCTACCAGCTTCTGGGCGCGGTATTGCCTCACG 3756

Qy     361 GGCAGGGTGGTCATTGTGGGTAGGGATCATNCTTGTCCGGGGGCCTATTGTTCCCGGAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3757 GGCACGGTGGTGCAGGGATCATCATCTTGTCCGGGGAGATCATCATTGTCCGGGGGCCTATCATCATTCCCGAC 3816

Qy     421 AGGGAGTCCTCTACCAGGGAGTTCGATGAGATGAGAGTGGAGGGTCGCGTCGCACCTTAC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3817 AGGGAAGGTCCTTTACCGGGAGTTCGAITGAGAGATGGAAGAGAGGTCGGCGTCGCGTCGCACCTCCTTAC 3876

Qy     481 ATCGAGCAGGGAAATGCAGCTCGAGTCGTCGCCGACAATTCAAAGACAAGGCCCTCGGGTTGCAG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3877 ATCGAACAGGGAAATGCAGCTCGAGTCGTCGCCGAACAATTCAAAAGCCCAATCGGGTTGCAA 3936

Qy     541 ACAGCCACCAAGCAAGCGAGGAGGCCGCTTCCCGTGGCTCCGGTTGCAAGTGGGCCCCTT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3937 ACAGCCACCAAGCAAGCGGAGGCTCGTGCTCCCGGTGGTGAAATCCAAGTGGGGGGACCTC 3996

Qy     601 GAGACTTCTGGGGCGGAAACACATGTGGGAACTTCATCAGCGGGATACAGTACTTAGCAGGC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3997 GAAGCCTTTCTGGGCGAAGCATATGTGGGAAATTTCATCAGCGGGATACAAATATTTAGCAGGC 4056

Qy     661 TTGTCCACTCTGCCTGGGGAAATCCCGGATTACCATCACTGATGGCATTCACAGCCCTCTGTC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4057 TTGTCCACTCTGCCTCGGGAAATCCCGGATTACCATCACTGATGGGCATTCACAGCCCTCTATC 4116

Qy     721 ACTAGCCCGTCACCACCCAATCTACCCTCCTTGCTTAACATCCTGGGGGTGAGGGCTAGCC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4117 ACCAGCCCGCTCACACCCAAACATACCCCGCGGATGGGGCATTCACAGCCCTGGCATAA 4176

Qy     781 GCCCAACTCGCTCCCCCAGTGGCTGCTTCAGCTTTCGTAGGCGCCGCCGCCAATTGCTGGCTGCG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4177 GCCCCAACTTGCTCTCCCAGGCTGCGCTTCTGCGCCTGGCGCCCGCCCCATGCGCTGGACGCG 4236

Qy     841 GCTCGTTGGCAGGCATAGGGCCTTGGGGAAGGTGGTCGGGTTGGTCCCCGGGCCTATGGGAGCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4237 GCTCGTTTGGCAGCATAGGCCTTGGGAAGGTGCTTGCCAGGTTATGGGAGCA 4296

Qy     901 GGAGTGGCCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGCGCGAAATGCCCTCCACCGAG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4297 GGGGTGGCCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGCGCGAAATGCCCTCCACCGAG 4356

Qy     961 GACCTGTTAACTTACTCTCCCGCCATCCTCTCTGGGTGGTGGGCCTGTCGGTGCGGGGTG 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4357 GACCTGGTTAACCTACTTCCCGTCGACTCGGTCGGCGCCGGCGGGGGGAGGGGGCTCGTG 4416

Qy    1021 TGCGCAGCGAATACTCCGTGTCGGCGGTCCAGGGGAGGGGGGCTGTTGTGCAGTGGATGAAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4417 TGCGCAGCGGATACTGGTCGGCACCGTGGGGCGCCCAGGGAGGGGGCTGCAGTGGGGATGAAC 4476
```

```
Db    6457 CTTATCGTGATTCCCAGATTTGGGGGGGTTCGTGTGTGTGCGAGAAAATGGCCCTTTACGGATGTGTG 6516
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3121 GTCTCCACCCCTCCCTCAGCGTGTGATGGGGCTCCTGGTACGGATTCCAGTATTCTCCTGGA 3180

Db    6517 GTCTCCACCCCTCCCTCAGCGCGTGATGGGCCTCTTCATACGGATTCCAATACTCTCCTGGA 6576
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3181 CACCGGGGTCGAGTTCCTTGTGAACGACGCCTTGGAAATCAAAGAGACCCCTATGGGGCTTTGCA 3240

Db    6577 CACGGGGTCGAGTTCCTTGGGTGAATCAGCGGGAAAAGGGAAGAAATGCCCTATGGGGCTTCGA 6636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3241 TATGACAACCCGCTCTGTTTTGACTCAACAGTCACTGAGAATGAACATCGTCGTAGAGGAGTCA 3300

Db    6637 TATGACACCCGCGTGTTTTGACTCAACGGTCACTGAGAATGACATCCGTCGGGCCTTACG 6696
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3301 ATTTATCAATGTTGTTGACTTGGCCCCGGAGGCCAGACCCATAAGGTGCTGCTCACAGAG 3360

Db    6697 CGGCTTTTACATCGGGGTGCCCCGGAGGCCAGACGCATAAGGTTCGGCTCACGGGTCTGCGTCACG 6756
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3361 CGGCTTTTATATCGGGGTCCCCTGGCCCCCGAAAGGGCCAGAACTGCGTTGGGCTATCGGCTCACG 3420

Db    6757 CGGCTTTACATCGGGGGGCCCCCTGCTAATTCTAAAGGGCCGAGACTGGGCTATCGGCTCACG 6816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3421 TGCCGCGGCGAGCGCGGTGCTGACGGGTAATACCCTCACATGTTACTTGAAG 3480

Db    6817 TGCCGCGCGGAGCGGGGTGTACTGACGGGTAATACCCTCACATGTTACTTGAAG 6876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3481 GCCTCTGCAGCCTGCTGCGAAAGCTTCCAGGACTGCACGATCTCGTCGTGTGGCCGGAGAC 3540

Db    6877 GCCCCTGCGCGCCGTTCGGAGCGTGGCGAAGCTTCCAGGACTGCACGATGCTCGTTTTATGCGGGAGAC 6936
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3541 GGGCCTTGTTCGTTTATCTGTGAGAGCGGCGGGGGAAACCCAGGACCCTACGGGCC 3600

Db    6937 GACCTTGTGTTTATCTGTGAAGCCCGGGACCCAAGACGACGAGGGGACCCTACGGGGCC 6996
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3601 TTCACGGAGGCTATGACTAGGTACTCTGCCCCCCCCCCCCCCCCCAACCAGAATAC 3660

Db    6997 TTCACGGAGGCTATGACTAGGATACTCTGCCCCCCCCCCCCCCCAAAACCAGAATAC 7056
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3661 GACCTGGAGTTGATAAACATCATGCTCCTCCATCCTCCCTCCAAATGTCCACAGGAATGCATCTGGC 3720

Db    7057 GACTTGGGAGTTGATAACATCATGTCTCCTCCTCCAAATGTGTCAGTGCGGGCCACGGATGCATCTGGC 7116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3721 AAAAGGGGTATACTACTCTCACCCCTGACCC 3749

Db    7117 AAAAGGGGTGTACTATCTCACCCGGTGACCC 7145
```

Qy 1081 ... 1140
Db 4477 ... 4536
Qy 1141 ... 1200
Db 4537 ... 4596
Qy 1201 ... 1260
Db 4597 ... 4656
Qy 1261 ... 1320
Db 4657 ... 4716
Qy 1321 ... 1380
Db 4717 ... 4776
Qy 1381 ... 1440
Db 4777 ... 4836
Qy 1441 ... 1500
Db 4837 ... 4896
Qy 1501 ... 1560
Db 4897 ... 4956
Qy 1561 ... 1620
Db 4957 ... 5016
Qy 1621 ... 1680
Db 5017 ... 5076
Qy 1681 ... 1740
Db 5077 ... 5136
Qy 1741 ... 1800
Db 5137 ... 5196
Qy 1801 ... 1860
Db 5197 ... 5256
Qy 1861 ... 1920
Db 5257 ... 5316
Qy 1921 ... 1980
Db 5317 ... 5376
Qy 1981 ... 2040
Db 5377 ... 5436
Qy 2041 ... 2100
Db 5437 ... 5496
Qy 2101 ... 2160
Db 5497 ... 5556
Qy 2161 ... 2220
Db 5557 ... 5616

Db 5557 ... 5616
Qy 2221 ... 2280
Db 5617 ... 5676
Qy 2281 ... 2340
Db 5677 ... 5736
Qy 2341 ... 2400
Db 5737 ... 5796
Qy 2401 ... 2460
Db 5797 ... 5856
Qy 2461 ... 2520
Db 5857 ... 5916
Qy 2521 ... 2580
Db 5917 ... 5976
Qy 2581 ... 2640
Db 5977 ... 6036
Qy 2641 ... 2700
Db 6037 ... 6096
Qy 2701 ... 2760
Db 6097 ... 6156
Qy 2761 ... 2820
Db 6157 ... 6216
Qy 2821 ... 2880
Db 6217 ... 6276
Qy 2881 ... 2940
Db 6277 ... 6336
Qy 2941 ... 3000
Db 6337 ... 6396
Qy 3001 ... 3060
Db 6397 ... 6456
Qy 3061 ... 3120
Db 6457 ... 6516
Qy 3121 ... 3180
Db 6517 ... 6576
Qy 3181 ... 3240
Db 6577 ... 6636
Qy 3241 ... 3300

RESULT 15

```
Db   6637  TATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA  6696
Qy   3301  ATTTATCAATGTTGTGACTTGGCCCCGGAAGCCAGACAGGGCCATAAGGTCGCCTCACGAG  3360
           |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6697  ATCTACCAATGTTGTGACTTGGCCCCGGAAGCCAGACAGGGCCATAAGGTCGCCTCACGAG  6756

Qy   3361  CGGGCTTTATCGGGGGGTCCCCTGACTAATTCAAAAGGGGCAGAACTGCGGCTATCGCCGG  3420
           |||||||| |||||||| ||||||||| |||||||||||||||||||||||||||||||||
Db   6757  CGGGCTTTACATCGGGGGCCCCCTGACTAAATTCAAAAGGGGCAGAACTGCGGCTATCGCCG  6816

Qy   3421  TGCCGGCGGCGGAGGCGGCTGCTGACGAGCTACAGGTCGCCGGTAATACCCTCACTTGAAG  3480
           |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db   6817  TGCCGGCGGCGGTGTACTGACGAGCTACAGGTCGCCGGTAATACCCTCACATGTTACTTGAAG  6876

Qy   3481  GCCTTCTGCAGCCTGTTCGCGAGCTGCAAGCTCCAGGACTGACAGATGCTCGTGTGCGGAGAC  3540
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6877  GCCCGCTGCGGCGCCTGTTCGCGAGCTGCGGAGCTTCCAGGACTGACAGATGCTCGTGTGCGGAGAC  6936

Qy   3541  GGCCTTGTCGTTATCTGGTGAAAGCGCGGGGGGACCCCCAAGAGGAGGACGCGGGGCCTACGAGTC  3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6937  GACCTTGTCGTTATCTGTGAAAGCGCGGGGGACCCCAAGAGGAGGACGCGGGGCCTACGAGTC  6996

Qy   3601  TTCACGGAGGCTATGACTAGGTACTCTGCGCCCCGGGGGGCCCGCCCCAACCAGAATAC  3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601  TTCACGGAGGCTATGACTAGGTACTCTGCGCCCCGGGGGGCCCGCCCCAACCAGAATAC  3660

Qy   6997  TTCACGGAGGCTATGACTAGATACTCTGCGCCCCGGGGGGACCCCCCCAAAACCAGAAATAC  7056
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db   6997  TTCACGGAGGCTATGACTAGATACTCTGCGCCCCGGGGGGACCCCCCCAAAACCAGAAATAC  7056

Qy   3721  AAAAGGGTATTACTACTACCTTCACCCGTGACCC  3749
           |||||||||||||||||||||||||||||||||
Db   7057  GACTTGGAGTTGATTAACATCATGCTTCCTCCAAATGTGTTCAGTTCGGCATGCTCTGGC  7116

Db   7117  AAAAGGGTGTTACTACTACCTTCACCCGTGACCC  7145
```

LOCUS   SSE242652   7989 bp   RNA   linear   SYN 11-MAY-2000
DEFINITION Hepatitis C virus replicon I377/NS3-3'UTR.
ACCESSION   AJ242652
VERSION   AJ242652.1 GI:5441834
KEYWORDS   core-neo gene; NS2 gene; NS3 gene; NS3
            proteinase/helicase; NS3/4A proteinase cofactor; NS4A
            gene; NS4b protein; NS5A gene; NS5A phosphoprotein; NS5b gene; NS5B
            RNA dependant RNA polymerase; polyprotein.
SOURCE   Hepatitis C virus replicon I377/NS3-3'UTR
  ORGANISM Hepatitis C virus replicon I377/NS3-3'UTR
            other sequences; artificial sequences; vectors.
REFERENCE   1
  AUTHORS   Lohmann,V., Korner,F., Koch,J., Herian,U., Theilmann,L. and
            Bartenschlager,R.
  TITLE   Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell
            line
  JOURNAL   Science 285 (5424), 110-113 (1999)
  MEDLINE   99321193
   PUBMED   10390360
REFERENCE   2 (bases 1 to 7989)
  AUTHORS   Bartenschlager,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-MAY-1999) Bartenschlager R., Institute for Virology,
            Johannes Gutenberg - University Mainz, Obere Zahlbacher Strasse 67,
            55131 Mainz, GERMANY
FEATURES   Location/Qualifiers
     source   1..7989
                     /organism="Hepatitis C virus replicon I377/NS3-3'UTR"
                     /mol_type="other RNA"
                     /db_xref="taxon:95363"
                     /focus
                     /note="Neomycin-selectable bicistronic subgenomic
                     hepatitis C virus replicon"
     source   join(1..341,1801..7758)
                     /organism="Hepatitis C virus"
                     /mol_type="other RNA"

```
                     /strain="1b"
                     /isolate="Con1"
                     /db_xref="taxon:11103"
                     /tissue_type="liver"
     source          1190..1800
                     /organism="Encephalomyocarditis virus"
                     /mol_type="other RNA"
                     /db_xref="taxon:12104"
     5'UTR           1..341
                     /evidence=experimental
     gene            342..1181
                     /gene="core-neo"
     CDS             342..1181
                     /gene="core-neo"
                     /function="selectable marker"
                     /codon_start=1
                     /product="core-neo fusion protein"
                     /protein_id="CAB46912.1"
                     /db_xref="GI:5441835"
                     /translation="MSTNPKPQRTKGRAMIEQDGLHAGSPAAWVERLRGYDWAQQTI
                     GCSDAAVFRLSAQGRPVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVTEA
                     GRDWLLLGEVPGQDLLSSHLAPAEKVSIMADAVRRLHTLDPATCPFDHQAKHRIERAR
                     TRMEAGLVDQDDLDEEHQGLAPAELFARLKARMPDGEDLVVTHGDACLGNIMVENGRF
                     SGFIDCGRLGVADRYQDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLD
                     EFF"
     misc_feature    1190..1800
                     /note="internal ribosome entry site (IRES)"
     CDS             1801..7758
                     /codon_start=1
                     /product="non-structural polyprotein"
                     /protein_id="CAB46913.1"
                     /db_xref="GI:5441836"
                     /translation="MAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSF
                     LATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS
                     SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVC
                     TRGVAKAVDFVPVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAA
                     YAAQGYKVLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLA
                     DGGCSGGAYDIIICDECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPH
                     PNIEEVALSSTGEIPFYGKAIPIEIIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAY
                     YRGLDVSVIPTSGDVIVVATDALMTGFTGDFDSVIDCNCVTQTVDFSLDPTFTIETT
                     TVPQDAVSRSQRRGRTGRGRMGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPA
                     ETSVRLRAYLNTPGLPVCQDHLEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQA
                     TVCARAQAPPPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEVTTHPITKYIMACM
                     SADLEVVTSTWVLVGGVLAALAAYCLTTGSVVIVGRIILSGKPAIIPDREVLYREFDE
                     MEECASHLPYIEQGMQLAEQFKQKAIGLLQTATKQAEAAAPVVESKWRTLEAFWAKHM
                     WNFISGIQYLAGLSTLPGNPAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPP
                     SAASAFVGAGIAGAAVGSIGLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVN
                     LLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESD
                     AAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQS
                     KLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNT
                     WHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCP
                     CQVPAPEFFTEVDGVRLHRFAPPCKPLLREESTFLVGLNQYLVGSQLPCEPEPDVAVL
                     TSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCTTRHDSPDADLIEA
                     NLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILRRSRKFPRAMPI
                     WARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRKRTVVLSESTVSSALA
                     ELATKTFGSSESSAVDSGTATASPDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLSDG
                     SWSTVSEEASEDVVCCSMSYTWTGALITPCAAEETKLPINALSNSLLRHHNLVYATTS
                     RSASLRQKKVTFDRLQVLDDHYRDVLKEMKAKASTVKAKLLSVEEACKLTPPHSARSK
                     FGYGAKDVRNLSSKAVNHIRSVWKDLLEDTETPIDTTIMAKNEVFCVQPEKGGRKPAR
                     LIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYGFQYSPGQRVEFLVNAWKAKKCPMG
                     FAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNC
                     GYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCTMLVCGDDLVVICESAGTQED
                     EASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTT
                     PLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIY
                     GACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARS
                     VRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFVAGYSGGDIYHSL
                     SRARPRWFWMCLLLSVGVGIYLLPNR"
     mat_peptide     1801..3696
                     /gene="NS3"
                     /product="NS3 proteinase/helicase"
     mat_peptide     3697..3858
                     /gene="NS4A"
                     /product="NS3/4A proteinase cofactor"
```

```
mat_peptide     3859..4641
                /gene="NS4B"
                /product="NS4B protein"
mat_peptide     4642..5982
                /gene="NS5A"
                /product="NS5A phosphoprotein"
mat_peptide     5983..7755
                /gene="NS5B"
                /product="NS5B RNA dependant RNA polymerase"
gene            1801..3696
                /gene="NS3"
gene            3697..3858
                /gene="NS4A"
gene            3859..4641
                /gene="NS4B"
gene            4642..5982
                /gene="NS5A"
gene            5983..7755
                /gene="NS5B"
3'UTR           7759..7989
                /gene="NS5B"
ORIGIN

Query Match           87.8%;   Score 3291.4;  DB 12;  Length 7989;
Best Local Similarity 92.4%;   Pred. No. 0;   Mismatches 286;  Indels 0;  Gaps 0;
Matches 3463;  Conservative 0;
```

```
Qy      1  TGGGAGGAGGGGTCTTCACAGGCCTCACCCACCACGTGGATGCCCACTTCCTGTCCCAAACAAAG    60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   3397  TGGGAGGAGGCTTCTTTACAGGCCTCACCCACACTAGACGCCACTTCCTGTCCCAGACTAAG   3456

Qy     61  CAGGCAGGAGGACAACTTCCCCTCACCTTGGTGGCGTACCAGCCTACTGTTGTCGGCGTAGGGCC   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3457  CAGGCAGGAGGAGAACAACTTCCCCTCACCTTGGTAGCATACCAGGCTACCGTGGGCGCCCAGGGCT   3516

Qy    121  CAGGCCCCACCTCCATCACATGGGAATCAAAATGTGGAAGTGTGTTCATCATACGGCGTAAAGCCTACT   180
           ||||| |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   3517  CAGGCTCCACCTCCATCGTGGGAACCAAAATGTGGAAGTGTCTTCATCATACGGCGTAAAGCCTACG   3576

Qy    181  CTGCGGCGGGGCCCAACACCCTTGCTGTGTTATAGGCTGGGGAGCCGTTCAAAAACGAGGGTCACCCTC   240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3577  CTGCACGGGCGGGCCAACGCCCCTGTCTGTATAGGCTGGGGAGCCGTTCAAAAACGAGGGTCACCCTCACG   3636

Qy    241  ACACACCCCATAACCAAAATTCATCATGGCCATGCAAGTGTCAGCGACCGTTGGCTCAACAA   300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3637  ACACACCCCATAACCAAAATACATCATGGCCATGCAAGTGTCAGCGACCGTTGGCTCAACAA   3696

Qy    301  AGCACCTGGGGTGCGTCGGTGGGCGGGGGTCCTTGCAGTCTCTGCTATTTGCTTTGACAACA   360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3697  AGCACCTGGGCTGTGTCGGTGGGCGGGGGTCCTTGCAGTCTCTGCTATTTGCTTTGACAACA   3756

Qy    361  GGCCAGGCGTGGTCATTCTGGGTAGGAGATCATCTTGTCCGGGAAAGCCGGCCATCATTCCCGGAC   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3757  GGCCAGGCGTGGTCATTGTGGGCGCAGGAATCATCTTGTCCGGGAAAGCCGGCCATCATTCCCGGAC   3816

Qy    421  AGGGAAGTCCTCTACCAGGAGTTCGATGAGTGGAAGAGAGTGGCTTCCCACCTCCCTTAC   480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3817  AGGGAAGTCCTTTACCGGGAGTTCGATGAGGTGGAAGAGAGTGGCGCGCCTCCCTTAC   3876

Qy    481  ATCGAACAGGGAAATGCAGCTTCGCCGAGCGTTCAAGCAAAAAGCGCTCGCGGTTGCTGCAG   540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3877  ATCGAACAGGGGAATGCAGCGTCGCCGAACCAATTCAAACAGAGGCCAATCGGGTTGCTGCAA   3936

Qy    541  ACAGCCACCAAGCAAGCGGAGCCGCGGCTGCCTCCGTGGTGGTGGGCTGGAGCCCTT   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3937  ACAGCCACCAAGCAAGCAGAGGCCAGGGCTGCCTCCGTGGTGGTGGGCTGGAGCCCTT   3996

Qy    601  GAGACCTTCTGGGCGGAAAACACATGTGGAACTTCATCAGTGACTACTTAGCAGGC   660
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3997  GAAGCCTTCTGGGCGGAAGCATAGTGGAATTTCATCAGTGATTAGCAGGC   4056

Qy    661  TTGTCCACTCTGCCTCGGGAATCCCCGGATTGCATCACTCACTGATGGGCATTCACAGCCCTCTGTC   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4057  TTGTCCACTCTGCCTCGGCAACCCCGGATAGCATCACTGATGGGCATTCACAGCCCTCTATC   4116
```

```
Qy    721  ACTAGCCCGCTCACCCAATCTACCCTCCTGCTTAACATCCTTGGGGTAGGGATGGGTTAGCC   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4117  ACCAGCCCGCTCACCCAACATACCCTCCTGTTTAACATCCTTGGGGGATGGGTGCC   4176

Qy    781  GCCCAACTCGCTCTCCCCCCGGGAGTGCTTCAGCTTTCGTAGGCGCCGCCATTGCTTGGTGCG   840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4177  GCCCAACTTGCTCTCCCCCACGCGTCTTCTCGTCTTAGGCGCCGCCATCCTGCTGAGCG   4236

Qy    841  GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTTGGGACAATCTTGGCGTGGGCCTAGGAGCA   900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4237  GCTGTTGGGCAGCATAGGCCTTGGGAAGGTGCTTGTTGGGGATATTTGGCGTGGGCCTAGGGCA   4296

Qy    901  GGAGTGGCAGGCGCGGCTCGTGGCCTTTAAGGTCATGAGCGCGGAAAATGCCCTTCCACGAG   960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4297  GGGGTGGCAGGCGCGGGCCTTTAAGGTCATGAGCGGCGGAGATGCCCTTCCACGGAG   4356

Qy    961  GACCTTGGTTAACTTAACTCCCTGCTCATCCCTGCTTTCCCCTGCCTTCGGGTCCTTCGGGTGTG   1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4357  GACCTGGTTAACCTACTCCCTCGCTATCCCTGCTTCCCCCTGGGTCGGGTTGGGGTCTTCGGGTG   4416

Qy   1021  TGCGCAGCGATACTGCGTCGGCACGTGGGTCCAGGGGAGGGGGCTGTGCAGTGGGGATGGAAC   1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4417  TGCGCAGGCGGATACTGCGGTCGGCACGTGGGTCCAGGGGGAGGGGGCTGTGCAGTGGGGATGAAC   4476

Qy   1081  CGGCTGGATAGCGGTTCGCCTCGCGCGGGGTAACCAGTGTTTCCCGCACCGGGGCTGTGCCTGAG   1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4477  CGGCTGGATAGCGGTTCGCCTCTTTGCGTTCCCCCACCGGGGCTGTGCCTGAG   4536

Qy   1141  AGCGACGCGGCCGCCGCAGCGTCACTCGAGATCCTCCCGGACATCTTACTCACTCCACTGTTG   1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4537  AGCGACGCTGCGCACGTGTCACTCGAGATCCTCTCTGTAGTGTTACCATCCACTCAGTGCTG   4596

Qy   1201  AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCACGCCCTGCTCTCGGTTCGTGGCTA   1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4597  AAGAGGCCTTCAACCAGTGGATCAACGAGGACTGCTCCATGCTGGTCGTTCCGGGTGGCTA   4656

Qy   1261  AGGGGATGTTTGGGACTGGATGATGCACAGTTTGGCTGACTTCAAGACCTGGCTCCAGTCC   1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4657  AGGGATGTTTGGGATGGATGCACGGGTGTTGACTGATTTCAAGACCTGGCTCCAGTCC   4716

Qy   1321  AAGCTCCTGCCGCGGGCGATTACGGGCGGCATCCCCTTTTCTCTAATGCCAACGTGGGTCGGCGGG   1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4717  AAGCTCCTGCCGCGGATTGCCGGGAGTCCCTTCTTCTCATCATCAACGTGGGTACAAGGGA   4776

Qy   1381  GTCTGGCGGGGGAGGGCATCATGCGGGAGTCGATCACCGGA   1440
           |||||||||||||||||||||||||||||||||||||||||
Db   4777  GTCTGGCGGGCGGGCGACGGGCCATCATGCGAAGCCAGATCACCGGA   4836

Qy   1441  CATGTCAAAAACGGTTCCATGAGGATCGTTGGCCCTAAGCCACTGTAGTAACATGTGGCAT   1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4837  CATGTGAAAAACGGTTCCATGAGGATCGTGGCCCTTAGTACACGTGGCAT   4896

Qy   1501  GGAACACATTCCCCATCAAACGCATACACCACCGGCGCCGCCCTCCCCGGCCAAAC   1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4897  GGAACATTCCCCATTAAGCCACACCACGGGCCCCGGCCAAAT   4956

Qy   1561  TATTCCAGGGCGCTGTGGCGGGTGGGCTGCTGAGCGGGGGCACGGGGTTCACGGGGG   1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4957  TATTCCTTAGGGGCGCCGTGGGGGGGTGGGCTGCTGAGTACGGGCGGGGTTCGGGGG   5016

Qy   1621  GATTTCCACTACGTGACGCAGCATGACCACTGACAAGGTAAAATGCCCGTGCCTGCCCA   1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5017  GATTTCCACTACGTGACGGGCATGACACTGACAAGTGTGCCCGTGCCTGC   5076

Qy   1681  GCCCCCGGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGTCCGCGGTGC   1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5077  GCCCCGGAATTCTTCACAGAAGTGGGGTGGCGGTTGCACAGGGTACGTCCGCGGTGC   5136

Qy   1741  AAACCTCTCCTACGGGAGGAGGAGGTCAACATTCAAGGTCGGGGCTCAACCAAATACCTGGTTGGG   1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5137  AAACCCCTCCTTACGGGAGGAGGAGGTCACATTTCCGGTGGGGCTCAACAATACTGGGTTGGG   5196
```

Qy   1801 TCGCAGCTTCCATGCGAGCCCGACCGGATGTAGCAGTGCTGCTCACTTCCATGCTCACCGAC 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5197 TCACAGCTCCCATGCGGAGCCCGACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 5256

Qy   1861 CCCTTCCCACATCACAGCAGGACGCTAAGCGCGCAGGCAGCAGTGCTCACTTCCCCCCTCC 1920
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5257 CCCTTCCCACATTACGGCGGGAGACGGCTAAGCGCGGGGGATCTCCCCTCC 5316

Qy   1921 TTGGCCAGCTCTTCAGCTAGGCAGTTGTTGCTGGCGGGTCGGCGGACATTACC 1980
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5317 TTGGCCAGTCATCAGCTTAGCCAGTGTCTGCGGCCTTCCTTGAAGGCAACGATCTCGC 5376

Qy   2041 GGGCGGGGACATTACCCGCGGTGGGAGTTCAGAGAGAACAAAGGTAGTAATCCTGGACTCTTTTCGAC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5437 GGGCGGGGACATCACCCGCGTGGGAGTTCAGAAAATAAGGTAGTAATTTGGACTCTTTTCGAG 5496

Qy   2101 CCGGCTCCGAGCGGGAGGAGGAGGAGGAGTGTCCGTCCCGGACGCGGAGAGTCCTGCGGGAAA 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5497 CCGGCTCCAAGCGGGAGGAGGAGGAGGAAGTATCCGTTCCGGGCGGGCCATGAGGAGG 5556

Qy   2161 CAAAATGACTTCCCCACCGGACCGTGACCTTCATCGAGGGCCAACTCCGGGGCCATGAGTG 2220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5377 CGTCATGACTCCCCCGGACCGTGACCTCATCGAGGGCGGCCATGAGTG 5436

Qy   2221 CTGGAGTCCTGGGAAGGCCACGGGCAACCGGGACTACGTCCCTCCCGGACTGGTTACATGGGGTCCCACTGGCCA 2280
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5617 TTAGAGTCCTGGGAAGGCCACGGGGACTACGTCCCTCCCAGTGCTGTCACGGGTGTCCCACTGGCCG 5676

Qy   2281 CCTACTAAGACCCCTCCTATTACCACCTCCACGGAGGAGGAAGCAGTTGTTCTGCACGAA 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5677 CCTGCCAAGGCCCCTCCGATACCACCTTCCACGCGGAGGAAGTATCCGTTTGTCCTGTGTCAGAA 5736

Qy   2341 TCCAACCGTGTCTTCTTGTCCCTGGCCGTTGCCACCAAAAGGCTTTTCGGGTAGCTCCGAACCG 2400
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5737 TCTACCGGTGTCTTCTTGGCCTTGGCGGAGCTGCCCACACCCTCCACTG 5796

Qy   2401 TCGGCCCGTCGACACGGGCACGGCCAACCGGGCCTCTCCTCCTGACCTTCGGACGACGGCGGA 2460
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5797 TCGGCCCGTCGACAGCGGCACGGGCAAACGGGCCCTCTCCGACCGACCCTCCGACTGACACGGCGAC 5856

Qy   2461 GCAGGATCTCACGTTGAGTGTGTATTCCTCCATGCCCCCCTTGAGGGGGAGCCCGGGGGAC 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5857 GCGGGATCCGACGTTGAGTCGGTACTCCTCCAATGCCCGGGGGGTGAGGAGGGACCCGGGGGAT 5916

Qy   2521 CCCGGATCTCAGCGACACGGGGTCTTGGTCTACGGTGAGGGCGGGTCGAGGAGGTCGTC 2580
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5917 CCCGGATTCCAGCGGACGGGGTCTTTGGTCTACGGTGAGGACCTCGTC 5976

Qy   2581 TGCTGCTCGGANTGCCTACACATGGACAGGGGCTCTGATCACGCGCCATGCGGCTGCGGGAA 2640
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5977 TGCTGCTCGATGTCCTTACACATGGACAGGGGGCCCTGATCACGCGCCATGCGGCTGCGGGAA 6036

Qy   2701 GCTACCACCATCCCGCCAGGCGCAAGCCGCGCGAGAGAAGAGTCACTTTGACAGCTGCGAA 2760
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6097 GCTACAACATCTCGCAGGCGCAAGCCTGCCGCGCAGAGAAGAGTCACTTTGACAGACTGCAG 6156

Qy   2761 ATCCTGGACGGATCACTCACTACCAGGAGGGTGCTCAAGAGAGATGAAGGCGAAGGCGGGTCACAGTT 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6157 GTCCTGGACGACGGACCACTACCGGGGACGTGCTCAAGGAGGATGAAAGGCGGAAGGGGTCCACAGTT 6216

Qy   2821 AAGGCTAAGCTTCTATCAGTGAGAGAGACAAGCTGACGCCCCCGCCAAA 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db   6217 AAGGCTAAACTTCTATCCGTGGAGGAAGCTGACGCCCCCACACATTCGGCCAGA 6276

Qy   2881 TCTAAAATTTGGCTATGGGGGCAACCTATCCAGCAAGGCCATTAACCAC 2940
            |||||||||||||||||||||||||||||||||||||||||||||||||

Db   6277 TCTAAATTTGGCTATGGGGGCAAAGGACAGTCCGGAACGCACCCGTTAACCAC 6336

Qy   3000 ATCCGGCTCCCGTGTGTGGGGACTTGTTGGGAGAGACACACTGAAAACCACACATC 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6337 ATCCGGCTCCCGTGTGGGGACTTGCTGGGAAGACACTGAAACCCAATTGACACCCATC 6396

Db   6397 ATGGCAAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGGAGGAGGCCGGCAAGCCAGCTCGC 6456

Qy   3061 CTTTATCGTGTTCCCAGACTTGGGGGTCCGTGTGTGTGGGGAGAAAATGCCCTTATGACGTG 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6457 CTTATCGTATTCCCAGATTTGGGGGGTTCGTGTGTGGGGAGAAAAAATGGGCCCTTTTACGATGTG 6516

Qy   3121 GTCTCCACCCTCCCCTCAGGCTGTGATCGTGATGGGCTCCTCGTACGCGGATTCTCTTCCTCTGGA 3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6517 GTCTCCAACCCTCCTCAGGGCGTGATGGGCTGTCTCATACGGGATTCCGAATACTCCTCTGGA 6576

Qy   3181 CAGCGGGGTCGAGTTCCTGGGTTGAAGCGCCTGGAAATCAAAGGAGACCCCTATGGGCTTTTGCA 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6577 CAGCGGGTCGAGTTCCTGGGTGAAATGCCCTGGAAAATGCCCTTATGGGCTTCGCA 6636

Qy   3241 TATGACAACCCCGCTGTTTGTGACTGAACAGTCACTGAGAATGACATCCGTGTGTAGGAGTCA 3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6637 TATGACACCCCGCTGTTTTGACTCAACGGGTCACTGAGAAATGACATCCGGTTTGGGAGGAGTCA 6696

Qy   3301 ATTTATCCAAATGTTGTGACTTGGCCTGCCCCGGAGCCAGCAGCCCATTAAGGGTGCCAGCAGAG 3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6697 ATCTACCAATGTTGTGACTTGGCCCCGGAAGCCAGCAGGCCAATAAGGGTGCCACAGAG 6756

Qy   3361 CGGCTTTATTCGGGGGGTCCCCTGACTAATTCAAAAAGGGCAGAAACTGCGGGCTATCGGCCGG 3420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6757 CGGCTTTACATCGGGGGGCCCCCTGACTAATTCTCACGGGCAGAAACTGCGGGCTATCGGCGG 6816

Qy   3421 TGCCGCGCGACCGGCGGCGTGCCTCACGACTAGCTGCGGTAATACCCTCACAGTTACTTGAAG 3480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6817 TGCCGCGGGACGGGCGTTTACTGACGACCAGTGCGGTAATACCCTCACAGTGTTACTTGAAG 6876

Qy   3481 GCCCTCTTGCAGCCCTGTCGAGCTGCAAAGCTCCAGGCACTGACGATGCTCGTGTTGTGCGGAGAC 3540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6877 GCGGCTCTGCGGCCTTGTGCAGCTGCGAAGCTCCAGGACTGCAGGATGTTCGGTATGCGGAGAC 6936

Qy   3541 GGCCTTGTCGTTATCTGTGAGAGGCGCGGGGAAACCCAGGAGGACGCGCGGAGCCTACGGAGTC 3600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6937 GACCTTGTCGTTATCTGTTGTGGAAAGCGCGGGGCAACCCAAGAGGACGCGAGGCCTACGGGGCC 6996

Qy   3601 TTCACGGACGGCTATAGTACTCTGCCCGCCCCCCGCCCCCCAACCAGAATAC 3660
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6997 TTCACGGGAGGGCTATAGACTAGGATACTGCCCGCCCCCCCCGCCCCCAACCAGAATAC 7056

Qy   3661 GACCTGGAGTTGATTAACATCATGCTCCTCCCAAATGTGTCGGTTCGGGCCACTCTGGC 3720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7057 GACTTGGAGTTTGATAACATCATGCTCCTCCCAAATGTGTCAGTGCGCGACGATGGACACTCTGGC 7116

Qy   3721 AAAAGGGTATACTACCCTCACCCGTGACCC 3749
            ||||||||||||||||||||||||||||||
Db   7117 AAAAGGGGTGTACTATCTCACCCGTGACCC 7145

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:　　　　February 21, 2005, 09:01:42　; Search time 1861 seconds
　　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　11928.549 Million cell updates/sec

Title:　　　　US-09-664-363-22
Perfect score:　3750
Sequence:　　1 TGGGAGGGGGGTCTTCACAGG..........ACTACTCACCGTGACCCG 3750
Scoring table:　IDENTITY_NUC
　　　　　　　　Gapop 10.0 , Gapext 1.0
Searched:　　4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:　　8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　　Maximum Match 100%
　　　　　　　　　listing first 45 summaries

Database :　　N Geneseq_16Dec04:*
　　　　　　　1: geneseqn1980s:*
　　　　　　　2: geneseqn1990s:*
　　　　　　　3: geneseqn2000s:*
　　　　　　　4: geneseqn2001as:*
　　　　　　　5: geneseqn2001bs:*
　　　　　　　6: geneseqn2002as:*
　　　　　　　7: geneseqn2002bs:*
　　　　　　　8: geneseqn2003as:*
　　　　　　　9: geneseqn2003bs:*
　　　　　　　10: geneseqn2003cs:*
　　　　　　　11: geneseqn2004as:*
　　　　　　　12: geneseqn2004aa:*
　　　　　　　13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Description
1	3746.8	99.9	3750	2	AAQ12241	Aaq12241 Encodes p
2	3292.2	87.8	8642	6	ABK88574	Abk88574 Hepatitis
3	3291.4	87.8	7987	6	AAD25321	Aad25321 Hepatitis
4	3291.4	87.8	7989	3	AAA98968	Aaa98968 Hepatitis
5	3291.4	87.8	7989	12	AAD25322	Aad25322 Hepatitis
6	3291.4	87.8	7992	6	ADJ57845	Adj57845 HCV repli
7	3291.4	87.8	7992	6	AAL47276	Aal47276 Hepatitis
8	3291.4	87.8	8001	3	AAA98967	Aaa98967 Hepatitis
9	3291.4	87.8	8637	6	AAA98966	Aaa98966 Hepatitis
10	3291.4	87.8	8638	6	ABK88595	Abk88595 Hepatitis
11	3291.4	87.8	8639	6	ABK88573	Abk88573 Hepatitis
12	3291.4	87.8	8649	3	AAA98969	Aaa98969 Hepatitis
13	3291.4	87.8	9605	6	ABK91424	Abk91424 Hepatitis
14	3291.4	87.8	9605	6	ABK91411	Abk91411 Hepatitis
15	3291.4	87.8	9605	6	ABK91425	Abk91425 Hepatitis
16	3291.4	87.8	9605	6	ABK91426	Abk91426 Hepatitis
17	3291.4	87.8	10690	6	ABK91412	Abk91412 Hepatitis
18	3291.4	87.8	10690	9	ACA61697	Aca61697 Hepetitis
19	3291.4	87.8	10690	10	ADC83762	Adc83762 pHCVNeo17
20	3291.4	87.8	10691	6	ABK91423	Abk91423 Hepatitis

21	3291.4	87.8	11076	3	AAA99965	Aaa99965 Hepatitis
22	3291.4	87.8	11313	12	ADP862264	Adp862264 Hepatitis
23	3291.4	87.8	11313	12	ADP862271	Adp862271 Hepatitis
24	3289.8	87.7	7141	6	AAD25333	Aad25333 Hepatitis
25	3289.8	87.7	7789	6	AAD25330	Aad25330 Hepatitis
26	3289.8	87.7	7987	6	AAD25329	Aad25329 Hepatitis
27	3289.8	87.7	7987	6	AAD25324	Aad25324 Hepatitis
28	3289.8	87.7	7989	6	AAD25326	Aad25326 Hepatitis
29	3289.8	87.7	7989	6	AAD25325	Aad25325 Hepatitis
30	3289.8	87.7	7989	10	ADD93734	Add93734 Hepatitis
31	3289.8	87.7	7991	6	AAL47279	Aal47279 Hepatitis
32	3289.8	87.7	7992	6	AAL47280	Aal47280 Hepatitis
33	3289.8	87.7	7992	6	AAL47277	Aal47277 Hepatitis
34	3289.8	87.7	7992	6	AAL47281	Aal47281 Hepatitis
35	3289.8	87.7	8732	10	ADE48121	Ade48121 RNA HCV r
36	3289.8	87.7	9605	6	ABK91431	Abk91431 Hepatitis
37	3289.8	87.7	9605	6	ABK91429	Abk91429 Hepatitis
38	3289.8	87.7	9605	6	ABK91432	Abk91432 Hepatitis
39	3289.8	87.7	9605	6	ABK91430	Abk91430 Hepatitis
40	3289.8	87.7	9605	6	ABK91428	Abk91428 Hepatitis
41	3289.8	87.7	9605	6	ABK91433	Abk91433 Hepatitis
42	3289.8	87.7	9605	6	AAD25332	Aad25332 Hepatitis
43	3289.8	87.7	9605	6	ABK91448	Abk91448 Hepatitis
44	3289.8	87.7	10690	6	ABK91435	Abk91435 Hepatitis
45	3289.8	87.7	10690	6	ABK91243	Abk91243 Hepatitis

ALIGNMENTS

RESULT 1
AAQ12241

ID　　AAQ12241 standard; cDNA; 3750 BP.
XX
AC　　AAQ12241;
XX
DT　　25-MAR-2003　(revised)
DT　　17-SEP-1991　(first entry)
XX
DE　　Encodes portion of PT-NANBH viral non-structural protein.
XX
KW　　post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.
XX
OS　　Non-A.
OS　　non-B hepatitis virus.
XX
PN　　GB2239245-A.
XX
PD　　26-JUN-1991.
XX
PF　　17-DEC-1990;　90GB-00027250.
XX
PR　　18-DEC-1989;　89GB-00028562.
PR　　27-FEB-1990;　90GB-00004414.
PR　　03-MAR-1990;　90GB-00004814.
XX
PA　　(WELL) WELLCOME FOUND LTD.
PA　　(HIGH/) HIGHFIELD P E.
XX
PI　　Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ;
XX
DR　　WPI; 1991-187584/26.
DR　　P-PSDB; AAR12599.
XX
PT　　Post-transfusional non-A non-B hepatitis polypeptide(e) - and also DNA
PT　　and antibodies used in diagnostic assays and in vaccines.
XX
PS　　Claim 10; Page 88-97; 108pp; English.
XX
CC　　This sequence probably encodes viral non-structural proteins of the PT-
CC　　NANBH viral genome which are antigenic. It was isolated from serum of
CC　　humans infected by the virus. See also AAQ12236-40 and AAQ12242. (Updated
CC　　on 25-MAR-2003 to correct PA field.)

XX

SQ   Sequence 3750 BP; 794 A; 1140 C; 1072 G; 744 T; 0 U; 0 Other;

Query Match   99.9%;   Score 3746.8;   DB 2;   Length 3750;
Best Local Similarity   99.9%;   Pred No. 0;   Mismatches 2;
Matches 3748;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

(Nucleotide sequence alignment of Query (Qy) and Database (Db) sequences, rows numbered from 1 to 2100, shown in paired Qy/Db alignment blocks.)

```
Qy    2101 CCGCTTCCGACGGGAGGAGGATGAGGGGGGAAGAGTGTCCGTTCCGGGCGGGGAGGATCCTCTGCGGGAAA 2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 CCGCTTCCGACGGGAGGAGGATGAGGGGGGAAGAGTGTCCGTTCCGGGCGGGGAGGATCCTCTGCGGGAAA 2160

Qy    2161 TCCAAGAAATTCCACCAGCGATGCCCGACGATTACAACCCTCGGCTG 2220
           |||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 TCCAAGAAATTCCACCAGCGATGCCCGACGATTACAACCCTCGGCTG 2220

Qy    2221 CTGGAGTCCTGGGAGGGCCCGGGACTTACGGTCCCTCCAGTGGTCCCACTGCCA 2280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CTGGAGTCCTGGGAGGGCCCGGGACTTACGGTCCCTCCAGTGGTCCCACTGCCA 2280

Qy    2281 CCTACTAAGACCCCTTCCTATACCACCTCCACGGGAGGAGGAAGAGGAGCAGTTGTTGTTCTGACAGAA 2340
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 CCTACTAAGACCCCTTCCTATACCACCTCCACGGGAGGAGGAAGAGGAGCAGTTGTTGTTCTGACAGAA 2340

Qy    2341 TCCACCGTGTCTTCTCTGCCCTGGGGGAGCTTGCCCACGGGAGGAGGGTACATTGGGTCCCACTGCCG 2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 TCCACCGTGTCTTCTCTGCCCTGGGGGAGCTTGCCCACGGGAGGAGGGTACATTGGGTCCCACTGCCG 2400

Qy    2401 TCGGCCGGTCGACACGGGCAACCGGGTCAACCTCCGGCGGGGAGGAGGGTAGTCTCCGAACCG 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 TCGGCCGGTCGACACGGGCAACCGGGTCAACCTCCGGCGGGGAGGAGGGTAGTCTCCGAACCG 2460

Qy    2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCCTTGAGGGGGAGCCCGGGGGGAC 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCCTTGAGGGGGAGCCCGGGGGGAC 2520

Qy    2521 CCCGGATCTCAGCGACGGGGTCTTGGTCTACCGTGAGTGAGGGGGTGAGGAGGACTCGTC 2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 CCCGGATCTCAGCGACGGGGTCTTGGTCTACCGTGAGTGAGGGGGTGAGGAGGACTCGTC 2580

Qy    2581 TGCTGCTCGATGTCTACACATGGACAGCGGCGCTCTGATCACGACACCCGGAGGAA 2640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 TGCTGCTCGATGTCTACACATGGACAGCGGCGCTCTGATCACGACACCCGGAGGAA 2640

Qy    2641 AGCAAGCTGCCCAATCAACGGGTTGAGCAACTCTTTGCTCGGTCACCACAACATGGTCTAC 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 AGCAAGCTGCCCAATCAACGGGTTGAGCAACTCTTTGCTCGGTCACCACAACATGGTCTAC 2700

Qy    2701 GCTACCACAATCCCCGCAGCGCGCAAGCCAGCGGAGAAGAGGTCACCTTTGACGACTGCAA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 GCTACCACAATCCCCGCAGCGCGCAAGCCAGCGGAGAAGAGGTCACCTTTGACGACTGCAA 2760

Qy    2761 ATCCTGGACGATCACTACCAGGAGGACGTGCTCAAGGAGGATGAAGGGGAAGAGCGGTCCAACAGTT 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 ATCCTGGACGATCACTACCAGGAGGACGTGCTCAAGGAGGATGAAGGGGAAGAGCGGTCCAACAGTT 2820

Qy    2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCGCCCCCACATTCGGCCAAA 2880
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCGCCCCCACATTCGGCCAAA 2880

Qy    2881 TCTAAATTTGGCTATGGGGCAAAGGAACGTCCGGAACCTATCCAGCAAGGCCCATTAACCAC 2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 TCTAAATTTGGCTATGGGGCAAAGGAACGTCCGGAACCTATCCAGCAAGGCCCATTAACCAC 2940

Qy    2941 ATCCGGCTCCGTGTGGGAGGACTTGTTTGGAAGACACACAAATTGACACACCACCATC 3000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 ATCCGGCTCCGTGTGGGAGGACTTGTTTGGAAGACACACAAATTGACACACCACCATC 3000

Qy    3001 ATGGCAAAAAATGAGGGTTTTCTGGGTCCAACCAGAGAGAGAGGAGGGCCGCAAGCCAGCTCGC 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 ATGGCAAAAAATGAGGGTTTTCTGGGTCCAACCAGAGAGAGAGGAGGGCCGCAAGCCAGCTCGC 3060

Qy    3061 CTTATCGTGTTTCCCAGACTTGGGGGGTCCCTTGTGTGTTGCGAGAAAAATGGCCCCTCTATGCGTG 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 CTTATCGTGTTTCCCAGACTTGGGGGGTCCCTTGTGTGTTGCGAGAAAAATGGCCCCTCTATGCGTG 3120

Qy    3121 GTCTCCACCCCTCCCTCAGGCTGTGATGGGCCTCCTCGTCTCCTGTACGGATTCCAGTGTATTCTCCTGGA 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 GTCTCCACCCCTCCCTCAGGCTGTGATGGGCCTCCTCGTCTCCTGTACGGATTCCAGTGTATTCTCCTGGA 3180

Qy    3181 CAGCGGGGTCGAGTTCCTTGGGTGAAACGCCTTGGAAATCAAAAGAAGACCCCTATGGGCTTTGCA 3240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 CAGCGGGGTCGAGTTCCTTGGGTGAAACGCCTTGGAAATCAAAAGAAGACCCCTATGGGCTTTGCA 3240

Qy    3241 TATGACACCCCCTGTTTTGACTCAACGAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3241 TATGACACCCCCTGTTTTGACTCAACGAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300

Qy    3301 ATTTATCAAATGTTGTGACTTGGCCCCCCCGAAGCCAGGACAGCAGGGCCATAAGGTCGCTCACAGAG 3360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3301 ATTTATCAAATGTTGTGACTTGGCCCCCCCGAAGCCAGGACAGCAGGGCCATAAGGTCGCTCACAGAG 3360

Qy    3361 CGGCTTTATTATCGGGGGGGTCCCCTGACTAATTCAAATCAAAGGGCAGAAGAGGCCTATCGCCCGG 3420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3361 CGGCTTTATTATCGGGGGGGTCCCCTGACTAATTCAAATCAAAGGGCAGAAGAGGCCTATCGCCCGG 3420

Qy    3421 TGCCGCGGCGGGACGGGGCCGGTCACTGGCCGTGCGGTAATACCCTACCACAGTTACTTGAAG 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3421 TGCCGCGGCGGGACGGGGCCGGTCACTGGCCGTGCGGTAATACCCTACCACAGTTACTTGAAG 3480

Qy    3481 GCCTCTGCAGCCTGTCGTGCTGAGCTGCAAAGTCCAGGACTGCACGGATGCTCGTCGTGTGCGGAGAC 3540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3481 GCCTCTTCCAGCCTGTCGTGCTGAGCTGCAAAGTCCAGGACTGCACGGATGCTCGTCGTGTGCGGAGAC 3540

Qy    3541 GGCCTTGTCGTTATCTGTTGTGACAGCGCGGGGGAACCCAGGAGGACCGAGGCCTACGAGTC 3600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3541 GGCCTTGTCGTTATCTGTTGTGACAGCGCGGGGGAACCCAGGAGGACCGAGGCCTACGAGTC 3600

Qy    3601 TTCACGGAGGCTATGACTAGGTTACTCTGCCCCGGGGGGGACGCCCAACCAGAATAC 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3601 TTCACGGAGGCTATGACTAGGTACTCTGCCCCGGGGGGGACGCCCAACCAGAATAC 3660

Qy    3661 GACCTGGAGTTGATAACATCGTCCTCCAAATGTGTCGCGGCGCGATGCATCTGGC 3720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3661 GACCTGGAGTTGATAACATCATCGTCCTCCAAATGTGTCGGCGCGATGCATCTGGC 3720

Qy    3721 AAAAGGGTATACTACCTCACCCGTGACCCG 3750
           ||||||||||||||||||||||||||||||
Db    3721 AAAAGGGTATACTACCTCACCCGTGACCCG 3750
```

```
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G, Pause A;
XX
PI   Kukolj G, Pause A;
DR   WPI; 2002-575382/61.
DR   P-PSDB; ABG30582.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 11; Page 59-69; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This sequence represents hepatitis C virus replicon S22-3, a
CC   self-replicating HCV polynucleotide molecule created from the replicon
CC   APGK12 (see ABK88753)
XX
SQ   Sequence 8642 BP; 1768 A; 2557 C; 2438 G; 1877 T; 0 U; 2 Other;

Query Match      87.8%;  Score 3292.2;  DB 6;  Length 8642;
Best Local Similarity   92.3%;
Matches 3462;  Conservative  2;  Mismatches 285;  Indels  0;  Gaps  0;
```

```
Qy      1 TGGGAGGGGGGTCTTCACAGGCCCTCACCCACGTGGATGCCCACTTCCACTTCCTGTCCCAAACAAAG     60
          |||||| || |||||  ||||||||||||||||||||| |||||||| || || | |
Db   4046 TGGGAGAGGCGTCTTTACAGGCCCTCACCCACATAGACGCCCATTTCTTGTCGGCTAGGGCC     4105

Qy     61 CAGGCAGGAGAGACAACTTCCCCTACCTGGTGGCGTACCAGGCTACTGTGTGGCGCCTAGGGCC    120
          ||||||||||||| |||||||||||||||| ||| |||||||| ||| |  ||| |||| ||||
Db   4106 CAGGCAGGAGAGCAACTTCCCCTACCTGGTAGCATACCAGGCCTACCGAGTGGTCGCCAGGGCT    4165

Qy    121 CAGGCCCCACCTCCATCAGGGGATCAAAATGTGGAAAGTTGTCTCATACGGGTAAAAGCCTACT    180
          ||||| | |||||||||  ||| | | ||||||||| ||| |||||||||||||    |||||
Db   4166 CAGGCTTCCACCTCCATCGTGGGAACCAAAATGTGGAAGAGTGTCTCATACGGGTCTTAAAGCCTAC    4225

Qy    181 CTGCACGGGGCCCAACCACCCCTTGCTCTGTATTAGGCTGGGAGGCCGTTCCATACGGGCTACAC    240
          |||||| || |||||  |||||  ||||  | ||| |||| || ||| ||| | |  ||||||
Db   4226 CTGGGACGGGGCCCAAATGTGGAAGAGTGTCTCATACGGGTGTTGACTTCAAGCCTAGG    4285

Qy    241 ACACACCCCATAACCAAAATTCATCAATGGCATGTCAGGCAACCTGGAGGTCGTCACG    300
          |||||| | || || ||||| | |||||||| |||| ||| |||||||| ||||| ||
Db   4286 ACACACCCCATAACCAAAATACATCATGGCATGGCCTGACCTGGTCGCTACG    4345

Qy    301 AGCACCTGGGTGCTGCTGGGTGGCGGGGTTCCTTGCAGCGTCCTGCGTATTGCTTGACAACA    360
          ||||||||||  | | |||| || || ||||||| ||| | | || | |||| |||||||||
Db   4346 AGCACCTGGGTGCTCGTGGTAGGCGGAGTCCTCGGCGCGTATTGCCTGACAACA    4405

Qy    361 GGCAGCCGTGGTCAGTTGGGGTAGGACTCATCTTTGTCCGGGAARGCCCGCCGCCCATTCCCGAC    420
          |||||||| |||| ||  ||  | | |||| | ||| |||| | || ||||||  |||||||||
Db   4406 GGCAGCCGTGCTCATTGTGTGGGCCAGGAATCATCTTTGTTGTTCATGGACATTCCCGAC    4465

Qy    421 AGGGAAAGTCCTACACGAGAGTTGAATGAGGAAGAGAGTCGGGTCGCCAACTCCCCTAC    480
          ||||| | | | || ||||| | ||| ||| || || |||  |  | |||||||||||
Db   4466 AGGGAAGTCCTTTACGGGAGTTCGAATGAGGATGAGAGAGTGCGGCCCTCACACCTCCCTTAC    4525

Qy    481 ATCGAGCAGGGAATGCAGTCTCGCCAGCAACAAAAAGCCTCGGCGGTTGCTGCTGCAG    540
          ||||||||||||  || ||| |||||| ||||| ||  || |||| || |||| |||
Db   4526 ATCGAACAGGAGAATGCAGTCTTGCCGGAACAATTCAAAACAGAATTCAAAACAGAGC    4585
```

```
Db   4526 ATCGAACAGGAGAATGCAGTCTTGCCGGAACAATTCAAAACAGAATTCAAAGTGGCAGCCCTT     4585

Qy    541 ACAGCCCACCAAGCAAGCGAGAGGCTGCCCTGCTCCCGGTGTGGAGTCCAAAGTGGCGAGCCCTT    600
          |||||| |||||| ||||| ||  || || | |||  | ||| ||| | |||||||||||  |
Db   4586 ACAGCCACCAGGAGAGCGGAGGCTCGCTGCCCTGCTCCCGGGGAATCCAAAGTGGCGAGCCCTC    4645

Qy    601 GAGACCTTCTGGGCGGAAAACATGTGGAACTTCATCAGCGGGATACAGTACTTAGCAGGC    660
          |||||||||| |||||||  ||||||||| ||||||||  | ||| | |||| |||||||
Db   4646 GAAGCCTTCTGTGGGGAAAGCGATATATGTGGAATTTCATCAGCGGGACAATTTAGCAGGC    4705

Qy    661 TTGTCCACTCTGCCTGGGAATCCCGCCGGATTGCATCACACTGATTGGCGTTCACAGCCTCTGTC    720
          ||||||||||||||||| ||||| |||| | ||| |||||||  |||| |||||| ||||  |
Db   4706 TTGTCCACTCTGCCTGGGCAAGCAATATGTGGAAATTTCATCAGCGGGATACAATATTTAGCAGG    4765

Qy    721 ACTAGCCCGCTTCACCCAAATTCTACCCTCCTGCTTAAACATCCTGGGGGGATGGGGTAGCC    780
          ||||||||||||||||| ||||| ||| | || ||||| | || | ||||||||||||||||
Db   4766 ACCAGCCCGCTTCACCCAAAATTACCCTTAGTGCTTAACACTGATGGGGGATGGGGTAGCC    4825

Qy    781 GCCCAAACTCGCTCCCCCGGTGTTCTCCAGTGTGCTTCAGCTTTCGTAGGCGCGCATTGCTGGTGCG    840
          ||||||| |||| |||||| | |  | ||||| |||| ||| ||| |||  |||| ||| ||| |
Db   4826 GCCCAACTTGTCCTCCCCAGTGTGCTGTCAGCTTTGGTAGGCGCGCATTCAGGGTGCGGGTGCG    4885

Qy    841 GCTGTTGGCAGCATAGGCCTTGGGAGGGGTGCTTGTTGGGACATCTTGGCGTGCGGGTCGGGAGCA    900
          |||||||| |||| ||||| |||| ||| ||||| ||||| ||| ||||| |  || | || |||
Db   4886 GCTGTTGGCAGCATAGGCCTTGGGAAGGTCGTCAATGTTGGGCAGTTATGCAGGTGCA    4945

Qy    901 GGAGTGGCAGGCGGCGCCCTCGTGGCGCCGGAAAATGCCCTTCCACGGAG    960
          |||||||| || | |||||||  |||| || | ||||| || |||  |||
Db   4946 GGGGTGGCAGGGCCTTCAGGGGCCCTTTAAGGTCATGAGCCGAAAATGCCCTCACGGAG    5005

Qy    961 GACCTGGTTAACTTACTCCCTGCCATCCTCCTGTCCCTGGTCCGTTCGGTGGGGGTCGTG   1020
          |||||||||||||||| |||||||  |||||||||| || |||| ||| || | |||||
Db   5006 GACCTGGTTAACCTACTCCCTGCTATCCTCCTGGTCCTGTGGGGCCGTGTCAGTGGGATCGTG    5065

Qy   1021 TGGCGAGCGGATTACTGCCGTCCCGGCACGTGGTCCCAGGGCGGGGAGGCTCGTGCAGTGGATGAAC   1080
          |||| || |||| ||| || |  |||||||||||||||| || || ||||| ||||| |||| |||
Db   5066 TGGCGACGGATTACTCGTCGGCACTGGTCCCCAGGGCGGGGGAGGCTCTTGCAGTGGATGAAC    5125

Qy   1081 CGGCTGATACGGTTGCCTCGGCCACCATTGCTGGGCGTGGGGTGTGCCAGAG   1140
          |||||||||||| |  ||  |||||||| | | | | ||| |||| |||| |
Db   5126 CGGCTGATAGGGTTCGCTTCGGCCACCCACGGCGACTATGTGCCAGAG    5185

Qy   1141 AGCGACGGCCGCAGCACGGTCGTCACTCAGATCCTCTCGCGGGTAAGATCTACCCAACTGTTG   1200
          ||||| ||| || || | |  | || |||||||||||||||| |||| | | || |||||||
Db   5186 AGCGACGGCTGCAGCACGGTGTCACTCAGATCCTCTCTAGTGTTATCACTCCAACTGCTG    5245

Qy   1201 AAGAGGCTTCACCAGTGGATTAACGGAGGACTGGCTCCAGCGCCCCGCCGTCGGGGCTA   1260
          |||| ||||||||  | |||||| || || ||| | || |||||| || ||| |||||
Db   5246 AAGAGGGCTTCACCAGTGGATCAACGAGGACTGCTCCAGCCATGCTCGGCGTCTCCGGCTA    5305

Qy   1261 AGAGATGTTTGGGACTGGGATAGGATTAACGGGTTGACTTCAACGACCTGGCTCCAGTCC   1320
          |||||||||||| |||||| ||| |||||| ||||| ||| || ||||| || ||||| |
Db   5306 AGAGATGTTTGGGACTGGGGATAATGCACAGTTTTGGTGACTTCAACGACCTGGCTCCAGTCC    5365

Qy   1321 AAGCTTCCTGCGCGGATTACCGGGAGTCCCCTTTTCTCAGTCCAACGTGGGTACAAGGGG   1380
          |||||||| ||||| |||||  ||||| |||| |||| |||||| ||||| | |||||||
Db   5366 AAGCTTCCTGCGCGGATTGCCGGGAGTCCCTTTTCTCAGTGCAAGTGGGTGTACAAGGGA    5425

Qy   1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCCCATGTGGAGCACAGATCACCGGA   1440
          ||| |||||| |||||| | |||| ||| || | | |||| |||| |||| ||||| |||
Db   5426 GTCTGGCGGGGAAGACTCATGCAGATCATGCAACTGCCCAGTGTGGAGCACAGATCACCGGA    5485

Qy   1441 CATGTCAAAAAACGGTTCCATGAGGATCGTTGGGGCCTAAGACCTGTAGTGAGCACATGTGGCAT   1500
          ||||| | |||| |||| ||| |||| |||| || | | |||||  |||| |  |||||||||||
Db   5486 CATGTGAAAAACTGTTCCATGAGGATCGTGGGGCCTAAGACCTGTAGTAACCACATGTGGCAT    5545

Qy   1501 GGAACAGCTCCCATCAATCAACGACATACACCACGCCCCTGGCGCAAAC   1560
          |||||| || || |||||||| |||| || ||||| |||| ||||||||
Db   5546 GGAACATTCCCCATTAACGGTACCACCACGGCCCCTGCGCAAAT    5605

Qy   1561 TATTCCAGGGCGCTGTTGGGGTGGGGTGCTGCTGAGGAGGTACGTGGGAGGGTACGTGGGGG   1620
          |||||||||||||||||| |||||||||||||| |||| || ||| || | | |||||||||
```

Db 5606 TATTCTTAGGGCGCTCTGGGCGGTGCTGCTGAGGAGTACCTGGAGGTTACCGCGGGGTGGGG 5665

Qy 1621 GATTTCCACTACGTGACGAGCTATGACGACTGACACAGTGAAAATGCCCGTGCCCAGGTTCCA 1680
Db 5666 GATTTCCACTACGTGACGGGCCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5725

Qy 1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGCGTGCACCAGTGACGCTTCCGGCGTGC 1740
Db 5726 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGCGTGCACCAGTGACGCTTCCGGCGTGC 5785

Qy 1741 AAACCTCTTCCTTACCGGGAGGAGGTCACATTCCAGGTTCGGTGGTTGGG 1800
Db 5786 AAACCCCTCTACCGGGAGGAGGTCACATTCCTGGTCGGTGGTTGGG 5845

Qy 1801 TCGGCAGCTCCCATGCCGAGCCCGAACCGGATGTAGCAGTGTCCATTCCATGCTCACCGAC 1860
Db 5846 TCACAGCTCCCATGGCGAGCCCGAACCGGACGTGCTCACTTCCATGCTCACCGAC 5905

Qy 1861 CCCTCCCACATCACAGCAGAGACGGCTAACGCAGCCTGCCCATTCCATGCTCACCGAC 1920
Db 5906 CCCTCCACATTACCGGCGGAGAGCGGGTCAGAAAATAAGGCTAGTGGC 5965

Qy 1921 TTGGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCTTCCCGGAAGGCGACAACATAGCCACC 1980
Db 5966 TTGGCCAGCTCATCAGCTACCCAGCTGTCTGTCGCGCTTCCTTGAAGGCGACAACATGCCACC 6025

Qy 1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGCCGCCCATGAGATG 2040
Db 6026 CGTCATGAGCTCCCCGGACCGCTGAACCTCATCGGCCGGCGGCAGGAGATG 6085

Qy 2041 GGCGGGGACATTACCCGCGTGGAGGTCAGAGAACCAAGGTAGTAGTAAATCCTGACTCTTTCGAC 2100
Db 6086 GGCGGGGAACATCACCCGCGTGGAGTCAGAAAATAAGGCTAGTGAATTTGCACTCTTTCGAG 6145

Qy 2101 CCGCTCCGAGCGAGCGGAGGAGGAGCGGGAAGTGTCCGTTCCCGGCGGGGAGATCCTGCGGAAA 2160
Db 6146 CCGCTCCAAGCGGAGGAGGAGGAAGTATCCGTTCCGGCGGGAGATCCTGCGGGAGG 6205

Qy 2161 TCCAAGAAATTCCCACCAGCGGATGCCGCCATGGGCCACGCCCGGATTACAACCCTCCGCTG 2220
Db 6206 TCCAGGAAAATTCCCTCGGAGGGCCATCATGGGCGACGCCCGGATTACAACCCTCCACTG 6265

Qy 2221 CTGGAGTCTCTGGGAAGGCCCCCGGGACTACGTCCCTCCAGTGTACAGCTCTTTCGCC 2280
Db 6266 TTRGAGTCTCCTGGAAGGACCCGGACTACGTCCCTCCAGTGTACACGGGTGTCCATTGCCG 6325

Qy 2281 CCTACTAAGACCCCTCCTATACCACTTCCACGGAGGAGAGACAGTTGTTCTGACAGAA 2340
Db 6326 CCTGCCAAGGACCCTCCGATACCACCTCCACGGAGGAGAGAGGAGACCGGTTGTCCGTGTCAGAA 6385

Qy 2341 TCCACCGTGTCTTCTGCCCTGCCCACGAGCTTGCCACGACCGGCTTTCGGTAGCTCCGAACCG 2400
Db 6386 TCTACCGGTGTTCTTGCCCTTGGGCCTCGGACTCGGCCACTATGGGCGCGGCGAGTCA 6445

Qy 2401 TCCGCCCGTGCACAGCGGCACGCGCGCCTCCTGTGACAAACCTCCGACGACAGGCGGGA 2460
Db 6446 TCGGCCGGTCGACACGGCGCACGGCAACGGGCTCTCCTGACCAGCGACGCGGAC 6505

Qy 2461 GCAGGAGCTGACGTTGAGTCGGTATTCCTCCAGGCCCCCCTTGAGGGGGGGGGAC 2520
Db 6506 GCGGGGGATCCGGACGTTGAGGTCGTACTCCTCCATGGCCCCCCTTGAGGGGAGCCGCGGGGAT 6565

Qy 2521 CCCGGATCTCAGCGACGGGGGGTCTTGGTCTACCGTGACTGAGGACAGCTCGTC 2580
Db 6566 CCCGGATCTCAGCGACGAGGGGTCTTGGTCTACCGGTGAGGAGACGTCGTC 6625

Qy 2581 TGCTTGCTTGACATGTCTACACAGTGACAGGCGGCCATCGCTCATTCGGCGCC 2640
Db 6626 TGCTGCTCGATGTCCTACACAGTGACAGGCAACGGCCCTGATCACGCCTGCGAGGAA 6685

Qy 2641 AGCAAGCTGCCCATCAAGCGGGTTGAGCAACTCTTTGCTGGGTCACCACAACATGGTCTAC 2700
Db 6686 ACCAAGCTGCCCATCAAGCGGGTTGAGCAACTCTTTGCTGGGTCACCACAACTTGGTCTAT 6745

Qy 2701 GCTACCCACATCCCGCACGCGCAAGCGCACGGCGCGCAAGCCCAGAGAGAAGGTCACCCTTTTGAGAGACTGCAA 2760
Db 6746 GCTACAACATCTCGCACGCGCAAGCCCTGCGCGGAGAAGAAGGTCACCTTTTGACAGACTCCAG 6805

Qy 2761 ATCCTGGACGATCACTACCAGGACACGTTGCTCAAGGAGAAGGATGAAGGGCAGGTCCACAGTT 2820
Db 6806 GTCCTGGACGACCACTACCGGGACGGTGCTCAAGGAGATGAAGGGCCGGTCCACAGTT 6865

Qy 2821 AAGGGCTAAGCTTCTATCAGTGAGTAGAGAAGACGGGGCTGCAACGTGGGAGGAAGTGC 2880
Db 6866 AAGGGCTAAACTTCTATCCGGTGGAAGGACCTGTTAAGGTGACGCCCCCACATTGCGGTCCACGA 6925

Qy 2881 TCTAAATTTGGCTACTGGGGGGCAAAGGGCGTCCGGGAAGCCTATCCCCAGCAAGGCCATTAACCAC 2940
Db 6926 TCTAAATTTTGGCTACTGGGGGCAAAGGACGTCCGGGAACTATCCCAGGCAAGCCCGTTAACCAC 6985

Qy 2941 ATCCGGCTCCGTGTGGGAGGGACTTGTTGGAAGGACAGCACTGAAAACACCAATTGACACCATC 3000
Db 6986 ATCCGGCTCCGTGTGGGAGGGACTTGCTCGGAAGGCACACCAATTGACACCACCATC 7045

Qy 3001 ATGGCAAAAAATGAGGTTTTTCTGCGGTCCAACCAGAGGAGAGGCAGACGCCAGCTCGC 3060
Db 7046 ATGGCAAAAATGAGGGTTTTCTGCGGTCCAACCAGAAGGCCGGCGCCCAAGCCAGTCGC 7105

Qy 3061 CTTATCGTGTTCCCAGACTTGGGTGGGTCGGTGTGTGTGCGGAGAAAAATGGCCCTCTATGACGTG 3120
Db 7106 CTTATCGTATTCCCAGATTTGGGGGGTTCGGTGTGTGGGAGAAAAATGGCCCTTTACAGTGTG 7165

Qy 3121 GTCTCCACCTCCTCAGGCGTTGATGACTTGGTTGGAAGCAGCAGAGCAGAGCATAAGGTCGGATTCCAGTATTCTGCTGGA 3180
Db 7166 GTCTCCACCCTCCCCAGGCGCGTGATAGGGCTCTTCATACGGATTCAAATACTCTCCTGGA 7225

Qy 3181 CAGCGCGGGTCGAGTTCTTCCTGGGTGAAACCCCTGACTAAATTCAAAAGAGAAGAACCCCCTATGGGGCTTTGCA 3240
Db 7226 CAGCGGGTCGAGTTCTTCCTGGGTGAAATGCCCTGACTAATTCTAAAGGGCAGAAACTGCGGGCTATCGCCGG 7285

Qy 3241 TATGACACCCGCCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGGTCGTGTTAGAGGAGTCA 3300
Db 7286 TATGACCGGTTCAACGGGTTGACTCACTGAGAATGACATCCGGTTGTTGGAGGAGTCA 7345

Qy 3301 ATTTATCAATGTTGTGACTTGGCCCCCCGGAAGCCAGAACACAGACCTAAGGTCGCGCTCACAGAG 3360
Db 7346 ATCTACCAAATGTTGTGACTTGGCCCCCCGAGAGCCAGAACACAGCCATAAGGTCGGCTCACAGAG 7405

Qy 3361 CGGGCTTTATATGCGGGGGTCCCCCTGACTAAATTCAAAAAGGGCAGAAACTGCGGGCTATCGCCGG 3420
Db 7406 CGGGCTTTACATCGGGGGCGGCCCCCTGACTAATTCTAAAGGGCAGAAACTGCGGGCTATCGCCGG 7465

Qy 3421 TGCCGGCGAGGCGGGTGCTGCAGGACTAGTCTGCGGTAATACCCTCACAGTGTTACTTGAAG 3480
Db 7466 TGCCCGGCGAGGCGGGTGTACTGACGCAGCCAAGCTGCGGTAATCACACAGTGTTACTTGGAG 7525

Qy 3481 GCCTCTGCACCTGTGTTGCAGAAGCTCAGGGACTCCAGGCTCGCAGGATGTGTGTGCGGAGAC 3540
Db 7526 GCCGCTGCGGCCGTGTGCGGAAGCTCCAGAGCTGCGCAGATGTTGTAGCCGGAGAC 7585

Qy 3541 GGGCTTGTCGTTTATCGTGTGAGAGGGGGGAACCCAGGGAACGCGCGCGGGGAGGAGTC 3600
Db 7586 GACCTTGTCGTTATCTCGTGTGGAAAGCGGGGGGCCCGCCAGGGAAGCGAGCGCCTACGGGCC 7645

Qy 3601 TTCACGGAGGGGGTATGACTAGGTACTCTGCGCCCCCCCGGGGGGGGCGCGCCAAACAGAATAC 3660
Db 7646 TTCACGGAGGGGCTATGACTAGGTACTGATACTCTGCGCCCCCCGCCAAGCGGAGCGCGCTTGGC 7705

Qy 3661 GACCTGGAGTTGATAACATCATCGCTCCTCCAAATGTTGTCGGTTGCGCCGCCACGATCGCCATCTGGC 3720
Db 7706 GACTTGGAGTTGTGATAACATCATGGCTCCTCCAAATGTTGTCAGTCGCGCGGATGCATCTGGC 7765

Qy 3721 AAAAGGGTATTACTTCACTCCCTCACCCGTGACCC 3749
Db 7766 AAAAGGGGTGTACTACTCTCACCCGTGACCC 7794

```
RESULT 3
AAD25321
ID   AAD25321 standard; cDNA; 7987 BP.
XX
AC   AAD25321;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Hepatitis C virus (HCV) replibBartMan/delta2U's cDNA.
XX
KW   Hepatitis C virus; HCV; transfection; infection; virus neutralisation;
KW   gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;
KW   ss.
XX
OS   Hepatitis C virus.
XX
PN   WO200189364-A2.
XX
PD   29-NOV-2001.
XX
PF   23-MAY-2001; 2001WO-US016822.
XX
PR   23-MAY-2000; 2000US-00576989.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Rice CM, Blight KJ;
XX
DR   WPI; 2002-066755/09.
XX
PS   Claim 44; Page 66-69; 174pp; English.
XX
CC   The invention relates to Hepatitis C virus (HCV) variants which include
CC   polynucleotides comprising non-naturally occurring HCV sequence and HCV
CC   variants that have a transfection efficiency and ability to survive
CC   subpassage greater than HCV that have wild-type polyprotein coding
CC   regions. The polynucleotides of the invention are useful for identifying
CC   a cell line that is permissive for infection with HCV and detecting
CC   replication of HCV in cells of the cell line. They are also useful for
CC   testing a compound for anti-viral properties and for inhibiting HCV
CC   infection. They are also useful for the generation of defined HCV virus
CC   stocks to develop in vitro and in vivo assays for virus neutralisation,
CC   attachment, penetration and entry, structure/function studies on HCV
CC   proteins and RNA elements and identification of new antiviral targets, a
CC   systematic survey of cell culture systems and conditions to identify
CC   those that support wild-type and variant HCV RNA replication and particle
CC   release, production of adaptive HCV variants capable of more efficiency
CC   replication in cell culture, production of HCV variants with altered
CC   tissue or species tropism, establishment of alternative animal models for
CC   inhibitor evaluation including those supporting HCV variant replication,
CC   development of cell-free HCV replication assays, production of
CC   immunogenic HCV particles for vaccination, engineering of attenuated HCV
CC   derivatives as possible vaccine candidates, engineering of attenuated or
CC   defective HCV derivatives for expression of heterologous gene products
CC   for gene therapy and vaccine applications and for utilisation of the HCV
CC   glycoproteins for targetted delivery of therapeutic agents to the liver
CC   or other cell types with appropriate receptors. Vaccine comprising these
CC   sequences is useful for inducing immunoprotection to HCV in a primate.
CC   The present sequence is Hepatitis C virus (HCV) replibBartMan/delta2U's
CC   cDNA.
XX
SQ   Sequence 7987 BP; 1647 A; 2368 C; 2243 G; 1729 T; 0 U; 0 Other;
```

```
Query Match             87.8%;  Score 3291.4;  DB 6;  Length 7987;
Best Local Similarity   92.4%;  Pred. No. 0;
Matches 3463;  Conservative 0;  Mismatches 286;  Indels 0;  Gaps 0;

Qy          1 TGGGAGGGGCGTCTTCACAGGCGCCTCACCCACTGTGGGATGCCCACTTCCTTGTCCAAACAAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3397 TGGGAGGAGCGTCTTTACAGGCGCCTCACCCACATAGACGCCCATTTCTTGTCCCGCTAGGCC 3456

Qy         61 CAGGGCAGGAGGAGACAACTTCCCCTACTACCTGGTGCGCGTACCCAGGGCTACGTGTCCGGGCTAGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3457 CAGGGCAGGAGGAGACAACTTCCCCTACTACCTGGTGCGCGTACCCAGGGCTACGTGTCCGGGCTACG 3516

Qy        121 CAGGGCCCACCTCCATCATGGGATCAAAATGTTGGGAAGTGTCTCATACGGGCTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3517 CAGGGCTTCCACCTTCCATCGTGGGACCAAAAGTGGAAGTGTCTCATACGGGCTACG 3576

Qy        181 CTGCGCCGGGGCCCAACACCCTTGCTTGTATAGGCCTGGAGCGCCTCCAAAAACGAGGTCACCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3577 CTGCACGGGCCCAACACCCTTGCTTGTATAGGCCTGTTCAAAACGAGGTTACTACC 3636

Qy        241 ACACACCCCCATAACCAAAATTCATCATGGCATGCATGTCAGGCCGACCTGGAGGTCGTCACG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3637 ACACACCCCATAACCAAAATACATCATGGCATCGATGTCGACCTGGAGGTCGTCACG 3696

Qy        301 AGCACCTTGGGTTGCTGGTGGGCGGGGTTCCTTGCAGCTCTTGCTTGGCTTTTGACACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3697 AGCACCTTGGGTGTCTGGTGGGGAGAGTTCCTAGCACTTGCTTGGCTGATTTGTGACAACA 3756

Qy        361 GGCAGGCGGTGGTTCATTGTGGGCAGGGATCATCTTGTGGGAAAGCCGGACGCGGCTATTGTTTCCGGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3757 GGCCAGGCGTGGTTCATTGTCCGGAGGATCATCTTGTGGGAAAGCCGGACGCGGCCATCATTCCGGAC 3816

Qy        421 AGGGAAGTCCTTCTACCAGGAGGAGTTCGATGAGGAAGCATGTCAGCGGCGCCGGGGTCGCCACCTCCTTAC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3817 AGGGAAGTCCTTTTACCAGGGAGGAGTTCGATGAGGAAGCATGTCGGCCGGGGTCGCCACACCTCCTTAC 3876

Qy        481 ATCGAGCAGGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAGCGTCTCGGGTTGCTGCAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3877 ATCGAACAGGGGAAATGCAGCTCGCCGAACAATTCAAAGAAAAAGCGCAATCGGGTTGCTGCAA 3936

Qy        541 ACAGCCACCAAGCAAGCGGGAGGCCCCTGCTCCCCGGGGTGGCAGGTGGCAGGCCCCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3937 ACAGCCACCAAGCAACGGGGAGGCTGCTGCTCCCCGGTGGTGGCAGGTGGCGGGGACCCTC 3996

Qy        601 GAGACCTTCTGGGCGAAACACATGTGGAACTTCATCATGAGGGAACTTCAGGGAGATACAGTACTTAGCAGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3997 GAAGCCTTCTGGGCGGAAGCATATGTGGAATTTCATCAGAGCGCCGGATACATATTTAGCAGGC 4056

Qy        661 TTGTCCACTCTGCCTGGGAATCCCCGGCGATTGCATCCATCGATGGCCGTTCACAGCCTCTGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4057 TTGTCCACTCTGCCTGGCAACCCGGCGATGAGCATCACTGATGGCCATTCACAGCCCTCTATC 4116

Qy        721 ACTAGCCCGCTCACCACACCCAATCATCACCTCTCGCTTAAACATCCTGGGGGGGATGGGTTAGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4117 ACCAGCCCGCTCACCACACATACCCTCGTTAAACATCCTGGGGGGGATGGGTGGCGGC 4176

Qy        781 GCCCAACTCGGTTCCCCCACAGTCGTGCTTCAGCTTCGTAGCTTCGTTGGTACGGCCCATTGCTGGTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4177 GCCCAACTTGCTCCTCTCCCAGGCGTTCTTCGTTGCTTTCGTTGGACATCTTGGCGGCATGCTGCTGGAGCG 4236

Qy        841 GCTGTTTGGGCAGCGATAGGCCTTGGGGAAGGTGCTTGTGTGGACACTTGCGCGGGCTTATGGAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4237 GCTGTTTGGGCAGCGATAGGCCTTGGGGAAGGTGCTTGTGTGGATATTTTGGCAGGGTTATGGAGCA 4296

Qy        901 GGAGTGGCAGCGCGCGCGCTCGTGCGGCCTTTAAGGTCATGAGCGGCGAAATGCCCTCCACGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4297 GGGGGTGGCAGGCGCGCTCGTGCGCCTTTAAGGTCATGAGCGCGAGGATGCCCTCCACGGAG 4356

Qy        961 GACCTGGTTAACTTACTCCCTGCCAATCCTCTCTCGTGGTCGTCGTGCCCTGGTTCGTCGTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4357 GACCTGGTTAACCTACTCCCTGCTATCCCTCTCCCCTGGGTCGTGCCCGTGGTTCGTCGTG 4416

Qy       1021 TGCGCACCGGATACTGCTCGGCGCACGTCGGTCCAGGGAGGCTCGCAGTGGGAATGAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4417 TGCGCAGCGGATACTGCGGTCGGCGACGTGGGCCCCAGGGGAGGGGCTCGTGCCAGTGGGATGAAC 4476

Qy       1081 CGGCTGATTAGGCGTTGCGGCTCGGCGACGTGGGTGTGGCTTCCCCACGCGCACGACTATGTGCCAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Qy 3301 ATTTATCAAATGTTGTGACTTGGTGGCCCCGAGCCAGACAGCAGGCCATTAAGGTCGGCTCACAGAG 3360
          || || |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db 6697 ATCTACCAAATGTTGTGACTTGGCCCCCGAGCCAGACAGGCCCATAAGGTTGGCTCACAGAG 6756

Qy 3361 CGGCCTTTATATCGGGGGTCCCCTACTAATTCAAAAGGCCAGAGACTGCGCGTATCGCCCGG 3420
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6757 CGGCGTTTACATCGGGGGCCCCCTGACTAATTCTAAAGGCCAGAACTGCGCGCTATCGCCCG 6816

Qy 3421 TGCGCGCGGGAGAGGCGGCGCTGAACGACTAGCTGCGGTAATACCTCACATGTTACTTGAAG 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6817 TGCGCGCGGCCAGCGGTGTACTGACCAACCCTCACACCCTCACATGTTACTTGAAG 6876

Qy 3481 GCCTTCTGCAGCCTGTCGAGCTGCAAAGTTCCAGGACTGCACGAGGCTTGGTTGTGCGGGAGAC 3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6877 GCCGCTGCGGCCCCTGTCGCGGCCTGCGGAAGCTTCCAGGACTGCACGAGGCTTGGTTATGCGGGGAC 6936

Qy 3541 GGCCCTTGTTGCGTTTACTGTTGAGAGCGGGCCGCGGGAGAACCCAGGAGGACGCGGAGGCCCTACGAGTC 3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6937 GACCTTGTCGTTATCTGTGAAAAGGCGGGGGGACCCAAGAGAGGACGAGGGCGGCCTACGGGGC 6996

Qy 3601 TTCACGGAGGCCTAIGACTAGGTACTCTGCGCCCCCGCCCCCCGGGGGCGCCCCAACCAGAATAC 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6997 TTCACGGAGGCTAGGACTAGATACTCTGCCCCCCCTGGGGAGGACCCCCCCAAAACCAGAATAC 7056

Qy 3661 GACCTGGAGTTGATAAACATCATGTCTCCCTCCAAATGTGTCGGTCGGGCGGCACCAGAATCTGGC 3720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 7057 GACTTGGAGTTGATAACATCATGCICTCTCCCAAAGTGTCAGTCGCGCGGCACGAATGCATTCTGGC 7116

Qy 3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
          ||||||||||||||||||||||||||||
Db 7117 AAAAGGGGTACTACTACCTCACCCGTGACCC 7145

RESULT 4

AAA98968

ID   AAA98968 standard; DNA; 7989 BP.
XX
AC   AAA98968;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Hepatitis C virus DNA fragment SEQ ID NO: 4.
XX
KW   Cell culture; therapy; infection; vaccine; diagnosis; gene therapy; ds.
XX
OS   Hepatitis C virus.
XX
PN   DE19915178-A1.
XX
PD   05-OCT-2000.
XX
PF   03-APR-1999;   99DE-01015178.
XX
PR   03-APR-1999;   99DE-01015178.
XX
PA   (UIMA-) UNIV MAINZ GUTENBERG JOHANNES.
XX
PI   Bartenschlager R;
XX
DR   WPI; 2000-629140/61.
XX
PT   Cell culture system for hepatitis C virus, useful e.g. in screening for
PT   therapeutic agents, comprises human hepatoma cells containing a viral RNA
PT   construct that includes a selectable gene.
XX
PS   Claim 8; Page 37-43; 58pp; German.
XX
CC   This invention describes a novel Hepatitis C virus (HCV) cell culture
CC   system comprising human hepatoma cells that contain an integrated HCV-RNA
CC   construct (I). (I) contains the HCV-specific RNA segments 5'-NTR (non-
CC   translated region), NS (non-structural)3, NS4A, NS4B, NS5A, NS5B and 3'-
CC   NTR, and a selectable (marker) gene (II). The cell cultures, and/or (I),
CC   are used to prepare, evaluate and/or test therapeutic and/or diagnostic
CC   agents for HCV infections, and to prepare vaccines against HCV infection
CC   (particularly preparation of attenuated HCV). The can also be used for
CC   preparation of a liver-specific delivery system for gene therapy, and to
CC   identify cells permissive for HCV replication. Virus RNA replicates
CC   autonomously and with high efficiency in this cellular system, so that
CC   variations in replication rates can be measured (for screening antiviral
CC   agents) quantitatively or qualitatively, using standard laboratory
CC   equipment. Efficient replication of HCV RNA is only achieved when the
CC   specified RNA segments are present and when the transfected cells are
CC   maintained under permanent selection pressure
XX
SQ   Sequence 7989 BP; 1647 A; 2368 C; 2243 G; 1731 T; 0 U; 0 Other;

Query Match   87.8%; Score 3291.4; DB 3; Length 7989;
Best Local Similarity   92.4%; Pred. No. 0;
Matches 3463; Conservative   0; Mismatches 286; Indels   0; Gaps   0;

Qy   1 TGGGGAGGGGGGTCTTCACAGCCCTCACCCACGTGGATGCCCACATTTCCTGTCCCAAACAAAG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3397 TGGGGAGGAGCGTCTTTTACAGGCCTCACCCACATAGACGCCCATTTCTTGTCCCAGACTAAG 3456

Qy   61 CAGGGCGAGGAGAACAGTTCCCCTACCTGGTGCGGTACCAGGCTTACTGTGTGCGGCTAGGGCC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3457 CAGGGCAGGGAGGACAACAGTTCCCCTACCTGGTAGCATTACCAGGCTTACGGTGTGCGGCT 3516

Qy   121 CAGGGCCCACCTCCATCATCGGGATCAAAATCTGGAAGTGTCTCATTACGGCCTAAAGCCTACT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3517 CAGGGTCCACCTCCATCGTGGGACCAAAATGTGGAAGTGTCCATACGGGCTAAAGCCTACG 3576

Qy   181 CTGCGCGGCGGGCCAACACCCTTTGCTGTTATAGGCTGGGAGGACCGCTCCAAAACGAGGGTCACCCTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3577 CTGCACGGGCGGCCAACCGCCCCGCTGTTATAGCCGTCGGGAGCGCCGTTCAAAACGAGGGTTACTACC 3636

Qy   241 ACACACCCCATAACCAAAATTCATCATGGGATGCATGCAGCCGACCTGGAGGTCGTCACG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3637 ACACACCCCATAACCAAAATACATCATGGCATGCATGTCCGCCTGACCTGGAGGTCGTCACG 3696

Qy   301 AGCACCTGGGTGCTTGGTCGGCGGGGGGGTCCTTGCAGCTCCTGGCTGCTGTTATTGCTTGACAACA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3697 AGCACCTGGGTCGGTGCTGGTAGGCGGAGTGCCTAGCAGTCTTGGCGCGCGGGTATTGCCTGACAACA 3756

Qy   361 GGCAGCGGTGGTCATTGTGGGTAGGATCATCATCTTGTCGCCGGGGGCGGCGGCTATTGTTCCCGAC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3757 GGCACGGTGGTCATTGTGGGCAGGANCATCTCTGTCGCCGGGGCCGCCGCCATCATTCCCGAC 3816

Qy   421 AGGGAAGTCCTCTACCAGGAGGTTCGATGAGATGGAAGGAGAGTGCGCCGTCGCACCTCCTTAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3817 AGGGAAGTCCTTTACGGGAGGTTCGATGAGGAAGAGAGTGCGCCGTCACACCTCCCTTAC 3876

Qy   481 ATCGAGCACGGGAATGCAGCTCGCCAGGAGCAGTTCAAGCAAAAAGCGCCTCGGGTTGCAG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3877 ATCGAACAGGGAATGCAGCTCGCCGAACAATTCAAACAGAAGGCAATCGGGTTGCTGCAA 3936

Qy   541 ACAGCCACCAAGCAAGCGCGAGGCGGCGCCCCTGCTCCCGGTGGTGGAGTCCAAGTCGGCGAGCCCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3937 ACAGCCACCAAGCCAAGCCGAGGGCTGCTCCCGGTGGTGGAATCCAAGTCGGCGGACCTC 3996

Qy   601 GAGACCCTTCTGGGCCGAAAACACATGTGGAACTTCATCAGCTGGTACTTAGCAGGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3997 GAAGCCCTTCTTCGGGCCGGAAGCATATGTGGAATTTCATCAGCGGGGAATACACATTTAGCCAGGC 4056

Qy   661 TTGTCCACTCTGCCTCGGAATCCCGCGGATTGCACTCGANTGGCGTTCACACCGCCTCTCGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4057 TTGTCCACTCTGCCTCGGCAACCCGCGGTAGCCATCGANTGGCCAATCAGACGCTTCTATC 4116

Qy   721 ACTAGCCGGCTTCACACCCAATCTACCCTCCTGCTTAACATCCTCTGGGGGATGGGGTAGCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4117 ACCAGGCGCGTCACCAACAATACCCTCCTGTTTAACATCCTCGGGGGATGGGGTCGGCC 4176

Qy   781 GCCCAACTCGCTCCCCCACAGTGCTTCAGCCTTTCGTAGGCGCCCGGAATGCGCATTGGTGGTGCG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4177 GCCCAACTTGCTCCTCCCAGGCGCCTTCGTAGGCGCCCGGCCATGCTGCTGGAGCG 4236

Qy 841 GCTCTTTGGGCAGCATAGGCCTTGGGAAGGTGCTTGGGGACATCTTTGGGCGGGGCTATGGGAGCA 900
        |||||||| ||||||||||||||| ||||||||||||||||| || ||||| ||||||||
Db 4237 GCTCGTTTGGGCAGCATAGGCCCTTGGGAAGGTGCTTGTTGGGATATTTTGGCAGGTTATGGGAGCA 4296

Qy 901 GGAGTGGGCAGGCGGCGCGGCTTCCTGGCCTTTAAGGTCATTGAGCGGGGCGAAATGCCCTCCACCGAG 960
        ||||||||||||||||||||| ||| ||||||| |||||||||||||||||||| ||||||||||||||
Db 4297 GGGGGTGGGCAGGCGGCGGCGCTTCGTGGCCTTTAAGGTCATGAGCGGGGCGAAATGCCCTCCACCGAG 4356

Qy 961 GACCTGGTTAACTTACTCCCTGCCATCTTCTTCCTGGTCGTTCGTCGGGGCGTGTG 1020
        |||||| | |||||||||||||||||||||||||||||||| ||||||||||||
Db 4357 GACCTGGTTAACCTACTCCCTGCTATCCTCTGCCTTGGTCGTTCGTCGGGGTCGTG 4416

Qy 1021 TGGCCAGCCGGATACTGCGTGGGCACGTGGGCCCAGGTGTGCAGTGGATGAAC 1080
        |||||||||||||| ||| |||||||||||||||||||||||||||||||||||
Db 4417 TGGCCAGCCGGATACTGCGTCGGCACGTGGGCCCAGGGGGGGGGGGGGGGGAAC 4476

Qy 1081 CGGCTGATAGGCGTTGCGCTTCGCGCGGGGTAACCATGTTTCCCCCACCGCCACTATGTGCCAGAG 1140
        ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db 4477 CGGCTGATAGCGTTCGCTTCGCGCGGGGTAACCACTCTCCCCCACCGCCACTATGTGCCAGAG 4536

Qy 1141 AGCGACCGCCGCAGCACGTGTTCACTCAGATCCTCTCCGACCTTACTTATCACCCAACTGTTG 1200
        |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db 4537 AGCGACGCGCTGCAGCACGTGTCACTCAGATCCTCTAGTTCTTACCATCACTCAGCTGTTG 4596

Qy 1201 AAGAGGGCTTCCACCAGTGGATTAACGAGACACTGCTCCACGCCCTCCTGCCTCTGGGCTA 1260
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4597 AAGAGGGCTTCACCAGTGGATCAACAGACTGCTCCACGGCCCTCCTGCCTCTGGGCTA 4656

Qy 1261 AGGGATGTTTGGGGACTGGATATGCACAGTTTGGGTGACTTCAAGACCTGGCTTCCAGTCC 1320
        |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db 4657 AGAGATGTTTGGGATTGGGATTATGCACGGTGTTGACTGATTTCAAGACCTGGTTCCAGTCC 4716

Qy 1321 AAGCTCCTCGCGGATTACCGGGGAGTCCCCTTTTTCTCATGGCCAACGTGGGTACAAGGGG 1380
        |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db 4717 AAGCTCCTGCGCCGGATTACCGGGGAGTCCCCTTCTTCTCATGTCAAGTCAACGTGGGTACAAGGGA 4776

Qy 1381 GTCTTGGGCGGGGAGACGGGCATCATGCGGACCACCTGCTCATGTGGGACGCACAGATCACCGGA 1440
        ||||||||||||||||| |||||||||||||||||||||||||||| |||||||||||||||||
Db 4777 GTCTGGGGCGGGACGGGCATCATGCAAAACACCTGTCCGGGCTCGTTGGCTA 4836

Qy 1441 CATGTCAAAAACGGTTCCATGAGATCGTTGGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
        ||||||||||||||||||||||||||| ||||||||||||| |||||||||||||||
Db 4837 CATGTGAAAAACGGTTCCATGAGACGTCATGAGGAGATCGTTAGTTAAACATGTGGCAT 4896

Qy 1501 GGAACATTCCCCAACTCAACCGCATACACCACCGGGGCCCCCTGCACGCGCCCAAAC 1560
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4897 GGAACATTCCCATTAACCGGTACACCACCAGGGCCCCCTGCAGCGGCGCCCAAAT 4956

Qy 1561 TATTCCAGGGCGCGTGTTGGGGGGTCGGCTGGGGGGTAGTGTAGTGGGGTTAGGGGGTTGGGG 1620
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4957 TATTCTAGGGCGCGTGTGGGGGTGGGCTGCTGCTGGAGGTTACGGCGGTGGGG 5016

Qy 1621 GATTTCCACTACGTGACGACCATGACATGACTGACAACGTAAAAATGCCCGTGTCCCAGGTTCCA 1680
        |||||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||
Db 4957 GATTTCCACTACGTGACGGGCCATGACCACTGACAACGTAAAGTGCCCGTGTCCAGGTTCCA 5076

Qy 1681 GCCCCGGAATTCTTCACAGAAGTGGATGGGGGTGCGGCTACGCTCCGCGCTGC 1740
        |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db 5077 GCCCCGGAATTCTTCACAGAAGTGGATGGGGGTGCGGGTTGGCACAGTTACGGCTCCACGCTGC 5136

Qy 1741 AAACCTCTCCTACGGGAGGAGGTCACACATTCCCAGCTCAACCAATACCTGGTTGGG 1800
        ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db 5137 AAACCCCTCCTACGGGAGGGTCACATTCCGGTCAACTCAAATACCTGGTTGGG 5196

Qy 1801 TCGGAGCTCCCATGGCAGCCCGGACCCGGATGTAGCAGTGCTTCACTTCCATGCTCACCGAC 1860
        ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 5197 TCACAGCTCCCCATGGCAGACCCGAACCCGGACGTAGCAGTGCTTCACTTCCATGCTCACCGAC 5256

Qy 1861 CCCTCCACACATCACAGAGAGACGGCTGCCAGGGGGTCTTCCCCCCTCC 1920
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 5257 CCCTCCACATTACGGCGGGAGACGGCTAGGCTGGCCAGGGGGATCTCCCCCCTCC 5316

Qy 1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTTGCTCGGCGCCTTCCTCGAAGGCGACGACATACATTACC 1980
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5317 TTGGCCAGCTCATCAGCTAGCCAGGCTGGCCGGCGCCTTCCTTGAAGGCAACATGCACTACC 5376

Db 5317 TTGGCCAGCTCATCAGCTAGCCAGGCTGGCCGGCGCCTTCCTTGAAGGCAACATGCACTACC 5376

Qy 1981 CAAAATGACTTCCCAGAGACTTCATCAGGCTCAACTCCTCTTGAGAGATG 2040
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 5377 CGTCATGACTCCCCCGGACGGCTGAACCTCATCGGGCAACCTTCCCGTGTGGGCGGGGAGATG 5436

Qy 2041 GGCGGGGGACACATTACCGGCGTGGGAGTCAGGAGAAACAGGTAGTTAGTCATCTTTCGAC 2100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5437 GGGCGGGAAACATCACCGGCGTGGGAGTCAGAAAATAAGGTAGTTCTTTCGAG 5496

Qy 2101 CCGCTCCGAGCGCGAGCGGAGGATGAGCGGGGGAAGTGTCGGTCCCGGCGCGGAAA 2160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5497 CCGCTTCCAAGCGGAGGGGAGGGAGGGGGAAGTATCGGTTCCGGGCGGGGAGAGG 5556

Qy 2161 TCCAAGAAATTCCCACCACAGGGATGTGCCGCATGGGAGGGGGGATTACAACCTCCGCTG 2220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5557 TCCAGGAAATTCCCGTCGGGTGGGGCGGGGAACGGGGGATTACAACCCTCCACTG 5616

Qy 2221 CTGGAGTCCTTGGGAAGGCCCCCGGGACTACAGTCCCCCAGTGTACATGGGGTGCCCACCTGCCA 2280
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5617 TTAGAGTGCTTGGAAGGACCCGGACCGACTACAGTACGGGGTGTCCCATTGCCG 5676

Qy 2281 CCTACTAAGACCCCTTCCTATACCACCTTCCACGGGAGGGGAGGAGGAGACAGTTGTTCTGACAGAA 2340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5677 CCTGCCAAGGCGCCCTTCCGGATACCACGTTGTTGAGTGGTTGCCTGTCAGAA 5736

Qy 2341 TCCACCGGTGTCTCTGCCCCGTGGGGGGGCGGGAGGCTTGCCACAAAGGGCTTTTCGGTTAGCTCCAACCG 2400
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5737 TCTACCGTCGTTCTCTGCCTTGGCGGGAGCTCGCCACACAGACCTTCGGCAGCTCGGAATCG 5796

Qy 2401 TCGGGCGGTCGACAGCGGCACGGGCCAACCGCCCCCTTCTCCGGACGACGGACGCGGA 2460
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5797 TCGGGCCGGTCGACACAGCGGACGGGCACCGGCTCTCTCCTGACCGACGAC 5856

Qy 2461 GCAGGGATCTGACGTTGAGTGTCGTATTCCTCCAGTCGGGGAGGAGGAGGAGGGGAGGAGGAC 2520
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5857 GCGGGGGATCGGACGTTGAGTCGTACTCTCCATGCCCCCTTGAGGAGGAGGACGGGGAT 5916

Qy 2521 CCCGATCTCAGCGAGACGGGGTCTTGGGTCTACGTGATGTGAGGGGCGGGAGCGGGAGCGGTGTC 2580
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5917 CCCGATCTCACGGACGGGTCTTTCCTTGGTCTCACTAAGGCAGGATGAGTGAGGACGGTCGTC 5976

Qy 2581 TGCTGCTTCGAGTGTCCTACACGGGACATGGGACCACATGGCCCATGGCCGGAGGGGAA 2640
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5977 TGCTGCTTCGATGTCCTACACATGCACGGGACCCGGATTCACGCGCGTGGGGAGGGA 6036

Qy 2641 AGCAAGCTGCCCATCACAACGCGTTGACCAACTCTTTGCTGCCCCATGGTCTTAC 2700
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6037 ACCAAGCTGCCCATCAATGCACTGACCAACTCTTTGCTGCCCACTGGTCTCTAT 6096

Qy 2701 GCTACACATCCCCGACGGGGGGCAGCGGCAAGCCAGAAGGTCACCTTTGACAGAGACTGCAA 2760
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6097 GCTACAACATCTCGCAGGGGACGCAAGGATCACCTTTGACAGAGACTGCAG 6156

Qy 2761 ATCCTGGACGGATCACTACCAGGACGTGCTCAAGGACGATGAAGGGGGAGGAGATCCCACAGTT 2820
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6157 GTCCTGGACGACACCACTACCGGGACGTTGCTCAAGGAGATGAAGGGGCGGTCCCACAGTT 6216

Qy 2821 AAGGCTAAGCTCTATCAGTGATAGAGAAGCCTGCAACATCAATTCGGCCAAA 2880
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6217 AAGGCTAAACTTCTATCCGTGGGAAAGAAGATCTGTAAGGCATTAACCAGA 6276

Qy 2881 TCTAAATTTGGCTATGGGGCAAAGAACGTCCGGACAGTCCGGAACCTATCCGCAGGCGGCCAC 2940
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6277 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGACATTAACCAC 6336

Qy 2941 ATCCGCGTTCCGTGTGGGGAGGACTTGTTGGAAGACACTGAAAACACCAAATTGACCACACCATC 3000
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6337 ATCCGCGTTCCGTGTGGGAAGGACTTGCTGTGGGAAGACACTGAGACACAAATTGACCACCATC 6396

Qy 3001 ATGGCAAAAATGAGGGTTTTCTGGGTCCAACCAGAGAGAGGAGGCCAAGGCAGCTCGC 3060
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   6397 ATGGCAAAAAATGAGGTTTTCTGGCTTCAACCAGAGAAGGGGGCGGCAAGCCAGCTCGC 6456
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3061 CTTATCGTGTTCCCAAGGACTTGGGGGTCCGTGGTGTGCGGAGAAATGGCCCTCTATGACGTG 3120
Db   6457 CTTATCGTATTCCCAGGATTTGGGGGTTCGTGTGTGCGGAGAAAATGGCCCTTTACGGATGTG 6516
                |||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3121 GTCTCCACCCTCCTCAGGCTGTCTCGTTGGATGGGGGCTCTTCCAGTATTCTCCCTGGA 3180
Db   6517 GTCTCCACCCTCCTCAGGCCGTGATTGGGGCTCTTCATACGGGATTCAATACTTCTCCTGGA 6576
                ||||||||||||||||||||||||||||||||||||||||
Qy   3181 CAGCGGGGTGGAGTTCCTGGTGAAGCGCCCTGGAAATCAAAGAGAGACCCCTATGGGGGCTTTGCA 3240
Db   6577 CAGCGGGGTGGAGTTCCTGGTGAATCCTGGAAAAGCGAAGAAAATGCCCTATGGGGGCTTTGCA 6636
                |||||||||||||||||||||||||||||||||||||||||||||
Qy   3241 TATGACACCCGTCTGTTTTGACTCACTGACTGAAGAAGATGACATCCGTGTTGTGACGGAGTCA 3300
Db   6637 TATGACACCCGCGTGTTTTGACTCAACGGTCACTGACGAATGACACCCGTGTTGTGACGGAGTCA 6696
                |||||||||||||||||||||||||||||||||||||||||||||
Qy   3301 ATTTTATCAATGTTGTGACTTGGGCCCCCGGAGCCAGACGGCCCATPAAGGTCGCTCACAGAG 3360
Db   6697 ATCTACCAAATGTTGTGACTTGGGCCCCCGGAAGCCAGACACGGCCCATPAAGGTCGCTCACAGAG 6756
                |||||||||||||||||||||||||||||||||||||||||||
Qy   3361 CGGGCTTTATATCGGGGGTCCCCTGACTAATTCTAAAAGGGCCAGAACTGCGGCTATCGCCG 3420
Db   6757 CGGGCTTTACATCGGGGGTCCCCTGACTAATTCTAAAAGGGCAGAACTGCGGCTATCGCCG 6816
                |||||||||||||||||||||||||||||||||||||||||
Qy   3421 TGCGGCGCGGGAGCGGGGGCGTGCTGACGACTAGCTGCGGGTAATCACCTTCACATGTTACTTGAAG 3480
Db   6817 TGCGGCGGGAGCGGGGTTGTACTGACAGCCAGTGCGTAATCCCTCACATGTTACTTGAAG 6876
                ||||||||||||||||||||||||||||||||||||||||
Qy   3481 GCCTCTGCAGCCCTGTCCGACTTGCAAAGCTCCAGGACTGCAGGATGCTCGTGTGTGCGGAGAC 3540
Db   6877 GCCGCTGCGGCCCTGTCCGACTTGCAAAGCTCAGGACTGCAGGATGCTCGTGTATGCGGAGAC 6936
                ||||||||||||||||||||||||||||||||||||||||||
Qy   3541 GGCCTTGTCGTTATCTGTGTGAGAGCGCGAGGGAGGACGCGGGGACGCCTACGAGTC 3600
Db   6937 GACTTTGTCGTTATCTGTGAAAGCGGGACCCAAGAGGACGAGGGACGCCTACGAGGCC 6996
                |||||||||||||||||||||||||||||||||||||||
Qy   3601 TTCACGGAGGCTATGACTAGTACTCTGCGCCCCGCCCCCGCCCCAACCAGAATAC 3660
Db   6997 TTCACGGAGGCTATGACTAGATACTCTGCCCCCTGCCCCCGCCCAAACCAGAATAC 7056
                ||||||||||||||||||||||||||||||||||||||||
Qy   3661 GACTGGAGTTGATANACATCATGCTCCTCCTCCAAATGTGTCGGTCGCCGCCACGATGCATCTGGC 3720
Db   7057 GACTTGGAGTTGATAACATCATGCTCCTCCTCAAATGTGTCAGTCGCCGCGATGCATCTGGC 7116
                ||||||||||||||||||||||||||||||||||||
Qy   3721 AAAAAGGGTATATACTTACCTCACCCGTGACCC 3749
Db   7117 AAAAAGGGTGTACTACTACTCACCGTGACCC 7145
```

FT   /note= "Nucleotide creating AvaII site"

```
Query Match             87.8%;  Score 3291.4;  DB 6;  Length 7989;
Best Local Similarity   92.4%;  Pred. No. 0;
Matches 3463;  Conservative  0;  Mismatches 286;  Indels  0;  Gaps  0;

Qy      1 TGGGAGGGCGTCTTCACGAGGGCCTCACCCAGGTGGATGGCCACTTCCTGTCCGTCCCAAACAAAG 60
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   3397 TGGGAGGAGCCGTCTTTACAGGCCCTCACCCACATGAGCGCCCATTTCTTGTCCGTCCCAGACTAAG 3456
Qy     61 CAGGCAGGAGCACAACTTCCCTACCTTGGTGCGTACCAGGTCACTGTGTGCGCTAGGGCC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3457 CAGGCAGGAGCACAACTTCCCTACCTTGGTAGCACATTACCAGGCGTGTGCGCTAGGGCT 3516
Qy    121 CAGGGCCCACTCCATTCCATCATGGGGATCAAAATGTGGAGAGTGTCTCATACGGGCTAAAGCCTACT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3517 CAGGGCTCCACCTCCATCCGTGGGACCAAAATGTTGGAGAAGTGTCTCATTACGGGCTACG 3576
Qy    181 CTGCGGGGGCCAAACACCCTTGCTGTPATAGGCTGGGGAGCCGTCCAAAAACGAGGTCACCCTC 240
            ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3577 CTGCACGGGGCCAACGCCCCTGCTTGTPATAGGCTGTPATAGGCTGGGAGCCGGTTCAAAACGAGGGTTACTACC 3636
```

```
Qy   5797  TCGGCCGTGGACACGGCGCAACGGGCCTCTCCTGACGACGGCGAC               5856

Qy   2461  GCAGGATCTGACGTTGAGTGTGTATTCCTCCATGCCCCCTTGAGGGGGGAGGCCGGGGGAC 2520
           || ||||| || || ||| |  | |||||||||||||||||||||||||||| |||||||||
Db   5857  GCGGGATCCGAGCTTGAGTCGTCACTCCTCCATGCCCCCTTGAGGGGGGAGGGCCGGGGGAT 5916

Qy   2521  CCCGATCTCAGCGGACGGGGTCTTGGTCTACCGTGAGTGAGGAGGCTCGGTC         2580
           |||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   5917  CCCGGATCTCAGCCGGACGGGGTCTTGGTCTACCGTGAGTGAGGAGGCTCGGTC        5976

Qy   2581  TGCTGCTCGAATGTCCTTACACATGGACACAGCGGGGCTCTGATCACGCGAGGAGGCGGAA 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5977  TGCTGCTCGATGTCCTACTGACATGCGACACAGCGGGGCTCTGATCACGCGGGGAGGCGGAA 6036

Qy   2641  AGCAAGCTGCCATCAAACGGTTGAGCAACTCTTTGCTCCGGTCACACAACATGGTCTAC    2700
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   6037  ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCCGGTCACACAACTTCGGTCTAT   6096

Qy   2701  GCTACCACCACTCCCCAGCCGCGCAAGCCCAAGCCGCGCAGAGAAGAGGTCACCTTTGACGACTGCAA 2760

Qy   2761  ATCCTTGGGACGATCACTACTACCAGGACGTGCTCAAGGAGATGAAGGCCAAGGCCTCACACAGTT   2820

Qy   2821  AAGGCTAAGCTTCTATCAGTAGTAGAGGAAGCCTGCAAGTGCCAACATTCGGCCCAA            2880

Qy   2941  ATCCGCCTTCCGTGTCGGGAGGACTTGTTGTGAAGACACCAATTGACACCACCATC            3000

Qy   3001  ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGAGGAGGCGCAAGCCAGCTCGC         3060

Qy   3061  CTTATCGTGTTCCCAGACTTGGGCGTCCGTGTCGTGTGGGAGGAGAAAAGTGGCCCGCTG        3120

Qy   3121  GTCTCCACCCTCCCTCAGGCCGTGATGGGCTCCGTGTACGGGTATTCTCGTTCTCCTGGA        3180

Qy   3241  TATGACACCCCGCTGTTTTGACTCACTGAGAATGACATCCGTGTAGAGAGAGTCA            3300

Qy   3301  ATTTATCAAATGTTGTTGGACTTGGCCCCGGAGCCAGGCCATAAGGTTCGCCTCACAGAG     3360

Qy   3421  TGCCCGCGGAGCGGGGCGTGACGACTAGCTGCGGGTTAATACCCTCACACATTACTTGAAG     3480

Qy   3481  GCCCTGCAGCCTGTTGCGAGCTGCACAGCTGCAAAGCTTCCAGGACTGCACGATGCTCTCGT... 3540
```

```
Qy   3541  GGGCCTTGTTCGTTATCTGTTGGAGGAGCGCGGCGGGGAAACCCAGGAGGAGCACGGCGGGGAGGCCTTACGGAGTC 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6937  GACCTTGTCTGTTACTGTGTGAAAGCCGCGGGGCGGGGAGCCCAAGAGGCAGACCTACGGGCC 6996

Qy   3601  TTCACGGAGGCTATGACTAGGTACTCTTGCCCCCCCGCGGGGCGACCGCCAGAATAC      3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6997  TTCACGGAGGCTATGACTAGATACTTCTGCCCCCGCCGGGGAGACCGGCCAGAATAC     7056

Qy   3661  GACCTGGAGTTGATAAACATCATGCTCCTCCAAATGTGTCGGTCGGCGACGATGCATCTGGC 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7057  GACTTGGAGTTGATAAACATCATGCTCCTCCAAATGTGTCGGCCACGGACGATGCATCTGGC 7116

Qy   3721  AAAAGGGTATACTACCTCACCCGTGACCC                                 3749
           |||||||||||||||||||||||||||||
Db   7117  AAAAGGGGTGTACTACCTCACCCGTGACCC                               7145
```

RESULT 6

ID   ADJ57845 standard; DNA; 7989 BP.
XX
AC   ADJ57845;
XX
DT   06-MAY-2004 (first entry)
XX
DE   HCV replicon encoding sequence.
XX
KW   hepatitis C virus; HCV; Antiinflammatory; Hepatotropic; Virucide; ds;
KW   HCV replicon.
XX
OS   Unidentified.
XX
FH   Key   Location/Qualifiers
FT   CDS   1801..7759
FT   /*tag= a
FT   /product= "HCV replicon"
XX
PN   WO200401513l-A2.
XX
PD   19-FEB-2004.
XX
PF   12-AUG-2003; 2003WO-US025260.
XX
PR   12-AUG-2002; 2002US-0402661P.
XX
PA   (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI   Gao M, Lemm JA, O'boyle DR, Nower P, Rigat K, Sun J;
XX
DR   WPI; 2004-180685/17.
DR   P-PSDB; ADJ57846.
XX
PT   Use of hepatitis C virus assays or reporter assays, e.g. identifying a
PT   compound that inhibits hepatitis C virus RNA replication or identifying a
PT   compound that modulates the activity of a gene of interest.
XX
PS   Claim 3; SEQ ID NO 1; 45pp; English.
XX
CC   The present invention relates to the use of hepatitis C virus (HCV)
CC   assays for identifying a compound that inhibits HCV RNA replication and
CC   reporter assays for identifying a compound that modulates the activity of
CC   a gene of interest. The assays are useful for identifying a compound that
CC   inhibits HCV RNA replication or for identifying a compound that modulates
CC   the activity of a gene of interest. The HCV assay is useful for high
CC   throughput screening that quantifies both the amount of HCV RNA
CC   replication inhibitory activity associated with a test compound and the
CC   amount of cytotoxicity associated with the test compound. The compound is
CC   useful for treating hepatitis C infection. Assays of the invention have
CC   distinct advantages when compared to QRT-PCR or other methods in that
CC   assays of the invention may take place in situ in a detergent based crude
CC   cell lysate, which requires no further preparation prior to performing
CC   the assays. The assays do not also involve numerous manipulations to add

or subtract reagents after addition of test compounds and are desirably based on a viral protein which is required by the HCV replicon for replication. The present sequence represents a HCV replicon encoding sequence used in the assay of the invention.

```
CC
CC
CC
XX
SQ   Sequence 7989 BP; 1647 A; 2368 C; 2243 G; 1731 T; 0 U; 0 Other;

Query Match         87.8%;  Score 3291.4;  DB 12;  Length 7989;
Best Local Similarity   92.4%;  Pred. No. 0;
Matches 3463;  Conservative 0;  Mismatches 286;  Indels 0;  Gaps 0;
```

```
Qy         1 TGGGAGGGGCGTCTTCACCAGCCCTCACCCACGTGGATGCCCACTTCCTGTCCCAAACAAAG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3397 TGGGAGGAGGGTCTTTACAGGGCCCTCACCCACCATAGAGCCCACTTTCTTGTCCCAGACTAAG 3456

Qy        61 CAGGCAGGAGACAACTTCCCCTCACCTGGGTGGGGTACCAGGCTACTGTGTCGCTAGGGCC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3457 CAGGCAGGAGGACAACTTCCCCTACCTGGTAGCATACCAGGCTACCAGGGCTCAGGGGCT 3516

Qy       121 CAGGCCCCACCTCCATCATGGGGGATCAAAATGTGGAAGTGTCTCATACGGGCTAAAGCCTACT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3517 CAGGCCTCCACCTCCATCGTGGGACCAAAATGTGGAAGTGTCTCATACGGGCTAAAGCCTACG 3576

Qy       181 CTGCCGCCGGCCCCAACACCCTTGCTGTATAGGCTGGGAGCCCTCCAAAACGAGGTCACCCTC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3577 CTGCACGGGCCCAACAGCCCCTGCTGTATAGGCTGGGAGCCCTTCAAAAACGAGGTTACTACC 3636

Qy       241 ACACACCCCATAACCAAAATTCATCATGCATGCAGCGGACACCTGATCGACCGGTCGTCACG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3637 ACACACCCCATAACCAAAATACATCATGGGCATGCAGCGGACACCTGGGAGGTCGTCACG 3696

Qy       301 AGCACCTGGGTGCTGGTGGTGGGGGGGGGGTTCTTGGCAGCTTCTGCTGTTATTTGCTTGACAACA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3697 AGCACCTGGGGTGCTGGTGGTCGTCCTAGCACGCTCTCGGCGGGTATTGCTCTGACAACA 3756

Qy       361 GGCAGGCGTGGTCATTGTGGGGTAGGAATCATCTTGTTCCGGCCGCGGCCCGGCTATTGTTCCGAC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3757 GGCAGGCGTGGTCATTGTGGGGCAGGATCATCATCTTGTTGTCCGGAAAGCCCGGCCATCATTCCGAC 3816

Qy       421 AGGGAAAGTTCCTCTACCAGAGAGTTGGATTGAGATGAGAGAGTGGGGGGTCGCCACCTCCCTTAC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3817 AGGGAAGTTCCTTTACCGGGAGTTCGATGAGATGGGAAGAGTTGGCTGCCGCTCACACCTCCCTTAC 3876

Qy       481 ATCGAGCAGGGAATTGCAGCTTCGCGCCCAGCAGTTCAAGCAAAAAGCGCTCGGGGTTGCTGCAG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3877 ATCGAACAGGGAATGCAGCTCGCGCCGAACAATTCAAAACAGAAGCAATCGGTTGCTGTGCAA 3936

Qy       541 ACACCCACCAAGCAAGCGAGCAGGAGCGCCCCTGCTCCCCGTGGTCGAGTGGCGAGCCCTT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3937 ACAGCCACCAAGCAACGCGGAGGGGCTCGCTGCTTCCCGTGGGTGGAATCCAAGTGGCGGGACCCTC 3996

Qy       601 GAGACCTTCTGGGCGAAACACATGTGGGAACTTCATCACGAGTACACTTAGCCAGGC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3997 GAACCCTTCTGGGCGAAAGCTATAGTGGGAATTCATCAGCGGGATACAAATATTTACGCAGC 4056

Qy       661 TTGTCCACTCTGCCTGGGAATCCCGCGATTGCATGGGGTTCATCATCAGGCCTCTGTC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4057 TTGTCCACTCTGCCTGCGCAACCCCGCGATTAGCATCACTGATGGCATTCACAGCCTTCTATC 4116

Qy       721 ACTAGCCCGCTCACCACCCAAATCTACCCTCCTGCTTAACATCTCTGGGGGGGATGGGGTAGCC 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4117 ACCAGCCCGCTCACCACCAACATACCCTCGTTTAAACATCCTGGGGTACTGGGGTGGGC 4176

Qy       781 GCCCAAACTCGGCTCCCCGGCTGCTGCAGCGCGCGCGGGGCAGTTGCTGGTTGCG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4177 GCCCAAACTTGCTCTCCTCCAGCGCGCTCCTTCGCTTCGTTAGGCGCGCGCATCGGCTGGAGCG 4236

Qy       841 GCTGTTGGGCAGCATAGGCCTTGGGAGAGGTCGTGGACATCTTGGCGCGGGCTATGGAGCA 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4237 GCTGTTGGCAGCATAGGCCTTGGGAAAGGTCGCTTGTGGGATATTTTGGGGCAGGTTATGGAGCA 4296

Qy       901 GGAGTGGCAGGCGGCCGCGCCTCGTGGGGGCGAAAATGCCCTCACCGGAG 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4297 CGTCATGAGACTCCCCGGACGCTGACCTCATCGAGGGCGGGCCAGAGAGATG 5436
```

```
Db      4297 GGGGTGGCAGGCGCGGCGGCTCGTGCGCCTTTAAGGTCATCGAGGCGCGAGATGAGCGCGGGGTCGTG 4356
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1021 TGGCCAGGCGGATACTGGGGTCGGCCCACGTGGGGTCAGGGGAGGAGGGGGCTGTGCAGTGGGATGAAC 1080

Qy       961 GACCTGGTTAAACTTACTCCCCTGCCATCCTCTCCTCCTTGGGTGCCCTTGGGGGTCGTG 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4357 GACCTGGTTAAACTACTACTCCCCTGCTATCCTCTCCCCTTGGGTCCCCTTAGTGTCGTCGGCGTG 4416

Qy      1081 CGGCTGGATAGGTTCGGGCTTCGGGGGGGGGGGTAACCCATGTTTTCCCCCACCGGGCCACTATGTGCCTGAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4417 TGGCGACGGATACTGCCTTCGGCGCACAGGTTGGGGCCCCCCCAGCGGAGGGGGGGGCTGTTGCCAGTGGGGATGAAC 4476

Qy      1201 AAGAGAGGCTTCACCCAGTGGGATCAACGGACGACTGTGCTCCACGGCCCATGTCCCCTCGCCTTGCCGTGGGG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4597 AAGAGGGCTTCACCCAGTGGATCAACGGAGGACTGCTCCAAGCGCCATGTCCGGGCTTCGCTTGGCTA 4656

Qy      1141 AGCGACGCCCGCAGCACGGCGTGTCACTCAGATCCTGGCCGAGGCCTTGCTCATCATGTCATCACCCAACTGTTG 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4537 AGCGACGGCTGCAGCGACACGCGTCAGCTGGATGCTCAGATCCTCTTAGTTCTTACCACATCACTCAGCGCTG 4596

Qy      1261 AGGGATGTTGGGGACTGGGAATATGCACAGTTTTGGCTTCAGAGCTTCAAGAGCCTTCCGAGTCC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4657 AGAGATGTTGGGATTGGGATAATGCAGCGGTGTTGATTGACTGAGTTTCAAGACCCTTCCAGTCC 4716

Qy      1321 AAGCTCCTGCCCGCGATTACCGGGAGTGGGGGGGGGGCCCCTTTTTTTTTTTTCATCCCAAGCAGTTGGGTACAAAGGGG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4717 AAGCTCCTGCGGGCCGGATTGCCGGGAGTCCCGGCGCGGTGTGGGGGCCCCCAGCGCCAAAAAT 4776

Qy      1381 GTCTGCGGGGAGACGGCATCATGGCGGGCATCATGCAAAACCACCTGCCCATGTGGAGGACGATCACCGGGA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4777 GTCTGCGGGGGCGACGCCATCAGCAAACACACTGCTTCATGTGGACACAGGGTACAAGGGA 4836

Qy      1441 CATGTCAAAAACGGTTCCATGAGGGATCGTCGTTGGGCCTTAAGACCTGTAGTAACATGTGGCAT 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4837 CATGTGAAAAAAAACGGTTCCATGAGGGATCCCTTTTTTTTTCATGGCAACCTTACTATCACGTTGCTTCACGTCC 4896

Qy      1501 GGAACATTCCCCATCAAACGACATACACCACGGGGGCCGGCCGCCCCTCCCAGCGCCCAAAC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4897 GGAACATTCCCCATTAACCGGTGGCTGCTACCACCACGGGGCGGGGGACCCTCCCAGCGCCCAAAAT 4956

Qy      1621 GATTTCCACTACGTGACGAGCAAGCATGACCACTGACAACGTAAAAATGCCCGTGGCCCAGGTTCCA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5017 GATTTCCACTACGTGACGCGGCCATGGCGCCCGGTCTAGGTAACGTGCAGGTTCCG 5076

Qy      1561 TATTCCAGGGGCGCCTGTGTGGCGGGCTGGGGGCGGGCATACGGCGGTGGGTGGCCGGA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4957 TATTCTAGGGGCCCTGTTGTGGGCGGGCTGGGGAGTGGTTTACGGGGTGGGTGGGG 5016

Qy      1681 GCCCCGAATTCTTCACAGAAGTGGAATGGGGGGGGGGGGTTGTCGCGCCAGCTGCC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5077 GCCCCCGAATTCTTCACAGACGGGTTGCGGTCTCCCCCTGCTCACTGGATCCCCTGC 5136

Qy      1741 TATTTCCAGGGGGCGTGGCGGGGGTCACATTCCTTGGTGGGGCCCCTTCAACAACAATACCTTGGTTGGG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5137 AAACCCCTCCTACGGGGAGGAGGTCACATTCCTGGTGGGCCTTCAATAACTACCTGGTTGGG 5196

Qy      1801 TCGCAGCTCCCATGCGAGCCCGGAACGGGATGTAGCAGTGCTCACTTCCATGCTCCACCGGAC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5197 TCACAGCTCCCCATGCGAGCCCGGAACGGGACTAGCAGTGCTCACTTCCATGCCTCACCGGAC 5256

Qy      1861 CCCTCCCACCATCACGACGAACGGCCTAGCGCAGCGGCTGGCCGCCCAGAGCTGGGGGGGTCTCCCCCTCC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5257 CCCTCCCACATTACGGCGGACGGCTAGGCTAGGCGAGGGATTCCCCCCCTCC 5316

Qy      1921 TTTGGGCAGCTCTTCAGCTAAGCCAGTTGTTCTGGCGCGCGCCTTCCTCGAAGGCGACACATTACC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5317 TTGGCAGCTCATCAGCTGGTCTCGCGCCTTCGTTGAAGGCGACAACATGCCACTACC 5376

Qy      1981 GGAGTGGCAGGCGCGGCCGCGCCTCGTGGGGCGCGAAAATGCCCTCACCGGAG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5377 CGTCATGAGACTCCCCGGACGCTGACCTCATCGAGGGCGGGCCAGAGAGATG 5436
```

[Page consists of a dense DNA sequence alignment, arranged in rotated paired Query (Qy) / Database (Db) rows with numeric coordinates and match bars (|). The alignment coordinates run from 2041 through 7145.]

```
XX
PR   07-NOV-2000; 2000US-0245866P.
XX
PA   (ANAD-) ANADYS PHARM INC.
XX
PI   Bichko V;
XX
PN   WPI; 2002-490082/52.
XX
DR   P-PSDB; AAO18000, AAO18001.
XX
PI
XX
PT   Novel nucleic acid encoding replication competent recombinant hepatitis C
PT   virus genome useful for screening anti-hepatitis C virus therapeutics and
PT   for vaccine development.
XX
PS   Claim 6; Page 43-47; 85pp; English.
XX
CC   The present invention provides protein and coding sequences from
CC   Hepatitis C virus (HCV), comprising all or part of the HCV genome and
CC   able to replicate efficiently when transfected into a susceptible cell
CC   line without reducing the growth rate of the cell line by more than 10
CC   fold. The sequences are useful for screening for anti-HCV therapeutics,
CC   for detecting antibodies to HCV in a biological sample such as blood,
CC   serum, plasma, blood cells, lymphocytes, or liver cells from a subject,
CC   for deriving authentic HCV components such as replication-component,
CC   infectious, replication-component, and replication-
CC   defective non-infectious infection-component, and replication-
CC   defective non-infectious HCV, in gene therapy or gene vaccination
CC   targeted to hepatic tissue for treating an animal infected or susceptible
CC   to HCV infection and for studying HCV infection and propagation. The
CC   present sequence is a clone of a fragment of the HCV genome which encodes
CC   the core-neo and NS3 proteinase/helicase proteins
XX
SQ   Sequence 7992 BP; 1648 A; 2369 C; 2243 G; 1732 T; 0 U; 0 Other;

Query Match             87.8%;  Score 3291.4;  DB 6;  Length 7992;
Best Local Similarity   92.4%;
Matches  3463;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;

Qy         1 TGGGAGAGGGGGTCTTCACAGGCCCTCACCCACCCACGTGGAGTGCCACTTGCCCAAACAAAG 60
             ||||||||||| ||||| |||||||||||||||| ||  ||  ||  ||  ||||||| |||||
Db      3397 TGGGAGAGGGGCTCTTTACAGGCCCTCACCCACATAGACGCCCATTTCTTGTCCCAGACTAAG 3456

Qy        61 CAGGCAGGAGACAACTTCCCCTACCTGGTGGCGTACCAGGCTACTGTGTGCGCTAGGGCC 120
             ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      3457 CAGGCAGGAGACAACTTCCCCTACCTGGTAGCATACCAGGCTACGGTGTGCGCCAGGGCT 3516

Qy       121 CAGGCCCCACCTCATCATGGGGATCAAAAATGTGAAAGTGGAAGTGTTCTCATACGGCTAAAAGCCTACT 180
             ||||| |||||||||||||||||||| |||||||| ||||||||||||||||||||||| ||||||||
Db      3517 CAGGCTCCACCTCCATCGTGGGACCAAAATGTGAAAGTGTTCATACGGGCTTAAAGCCTACG 3576

Qy       181 CTGCGCGGCGGCCAACACCCCTTGCTCTGTATAGGCTGGGAGCCGTTCAAAACGAGGTCACCCTC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3577 CTCGCACGGCGGCCAACGCCCCTCGCTGTATAGGCTGGGAGCCGTTCAAAACGAGGTTACTACC 3636

Qy       241 ACACACCCCATACCAAAATTCATCATGGCATGCAATGTCAGCGACCTGGAGGTCGTCACG 300
             ||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      3637 ACACACCCCATAACCAAAATACATCATGGCATGCATGTCGGCGAGGTGGTCAACACA 3696

Qy       301 AGCAACCTGGGTGCTGGTGGGGGCGGGGGTCCTTGCGACCTCTGCGCTCGGGGTATTGGTGGGCGGGA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3697 AGCACCCTGGGTGCTGGTAGGCGGAGTCTCTGCGCCGCTAGCGCTACGGCTGGGA 3756

Qy       361 GGCAGCGTGGTCATTGTGGGGTAGGATCATCTTGTTGTCCGGCGGGGCGGGGGGCGGGAC 420
             |||||||||||||||||||||||| |||||||||||| ||||||||||||| |||||||
Db      3757 GGCAGGCGTGGTCATTGTGGGCAGGATCATCTTGTTCCGGAAAGCCCGGGCGGGGGTGGGGAC 3816

Qy       421 AGGGAAGTCCTCTACCACAGAGTTCGATGAGATGAAAGAGTGCGGCCTCACGCTGCGG 480
             ||||||||||||||||||||| || |||||||||||||||||||||||||||||||||
Db      3817 AGGGAAGTCCTTTACCGGGGAGTTCGATGAGATGAGAGAGTGCGGCGCCTCACACCTTAC 3876

Qy       481 ATCGAGCAGGGAATGCACCGACTTCGGGCCGAGCAGTTCAAGCAAAAAAAGGGCTCGGGGTTGCTGCAG 540
             |||||||||||||||||||||||||||||||||| |||| ||||| ||||||||||||||
Db      3877 ATCAACAGGGAATGCAGCTCGCCGAGCAACAATTCAAAACACGAGGCAATCGGGTTGCTGCAA 3936

Qy       541 ACAGGCACCAAGCAACGGGAGGCCGCGGTTGCTCCCGGTGGAGTCCAAGTGGCGAGGCCCTT 600
             |||||||||||||| |||||||||||||||||||| |||||||||| ||||||||| |||
Db      3937 ACAGGCACCAAAGCAAGCGGAGGCCGCGTGCTTCCCCGGGTGGAGTTCCAAGTGGGGGAGCCTC 3996

Qy       601 GAGACCTTCTGGGGCGGAGAAACACATGTGGAACTTCATCAGCGGGGATACAGTAGTTAGCAGGC 660
             |||||| |||||||||||| |||||||||||||||||||||| |||||||||| ||||||||||
Db      3997 GAAGCCTTCTGTGGCGGAAGCATATGTGGAATTTCATCAGGGGGGATACAAGTAGTGAGGC 4056

Qy       661 TTGTCCACTCTTGCCTGGGAAATCCCGGGGAATTGCATCAGTGATGGGCGTTCACAGCCTGTC 720
             ||||| |||||||| |||||||| |||||||||||| ||||| |||||||| ||||||||||
Db      4057 TTGTCCACTCGTTGCCTGGGAAATCCCGGGGATAAGGGCATTCACAGCCTCATC 4116

Qy       721 ACTAGACCCGGCTCCACCACCCAATTCTACCCTCGTTTAAACATCCCTGGGGGGATGGGTAGCC 780
             |||||||||||||||||||||||||| |||||||| |||||||| |||||||||||||||| ||
Db      4117 ACCAGCCCCGGCTCACCACCCAACATCTACCCTCTGTTAAACATTCCTGGGGGGATGGGTAGCC 4176

Qy       781 GCCCAACTGCGTCCCCCAGTTGCTGCTTCAGCTTTCGGTAGGGCATTGCTTGGTGCG 840
             ||||||||||||||||||||||||||||| |||||||||| |||||||||||||||||
Db      4177 GCCCAACTTGCTCCTCCCAGGCGTGCTTGTTCTGCTTCGGTAGGTAGGCCATCGCTGGAGCG 4236

Qy       841 GCTGGTTGGCAGCATAGGCCCTTGGGAAGGTGCTTGGTGGACATCTTGGGCGTGGCCTGAGCA 900
             |||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||||||||
Db      4237 GCTGTTGGCAGCCATCAGGCCCTTGGGAAGGTGCTTGTGTGGATTATTTTGGCAGTTATGGAGCA 4296

Qy       901 GGAGTGGCAGCCGCGGCCGCTGCGTTTAAGGTCATGAGCGGGGCGAAATGCCCTCCACGGAG 960
             ||||||||||||||||||||| ||| |||| ||| ||||||||||||||| |||||||||||
Db      4297 GGGGGTGGCAGGCGCGGGCCTTAAGGTGCATGGAGCGGCGAGAATGCCCTCCACCGAG 4356

Qy       961 GACCTGGTTAACTTACTCCCCGTCATCCCTCCTCTCCGGGTCGCCCCTGTACTCGTCGTCGGGGGGTG 1020
             |||||||||| |||||||||||| ||||||||||| |||||||| ||||||||| |||||||||
Db      4357 GACCTGGTTAACCTACTCCCTGCTCATCCCTCCTCTCCTGGGTCGCCCTAGTCGTCGTGGGGGGTG 4416

Qy      1021 TGCGCAGGCGGATGGTACTGCGGTCGGCACGTGGGGTCCAGGGGGAGGGGGAGGGGATGAAC 1080
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||
Db      4417 TGCGCAGGCGGATAGTACTGGGTGGGCCACGTGGGGCCTGTGCAGTGGGATGGAAC 4476

Qy      1081 CGGCTCGATAGCGTTGCCTTGGCCCTCGCCGGGGTAACCATGTTTCCCCGCCACTAGTTGCCTGAG 1140
             ||||||||| |||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||
Db      4477 CGGCTGATAGGGTTCCGCTTCGCCGGGCTAACGGTCTCCCCACGGGTAACCATGTCGTCCTGAG 4536

Qy      1141 AGCGACGCCGCCGAGCAGCGTGTCACTCAGATCCTCTCCGGACTTGAGCATCCACCAACTGTTG 1200
             ||||||||||||||||||||||| ||||||| |||||| ||||||||||||||||||||||||
Db      4537 AGCGACGCCGCCGAGCACGTGTCAGTCGGATCCCTCTTAGTGTCTACCATCAGTAGTCAGCTGCTG 4596

Qy      1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCAGGTCTTCAGTGTGGGTGTGGCTA 1260
             ||||||||||||||||||||||||||||||||||| ||||||||||| |||||||||
Db      4597 AAGAGGCTTCACCAGTGGACGCCAACGACGGAGCTTCCAGGTCGTCGGTGGGTGTGGCTA 4656

Qy      1261 AGGGATGTTTGGGGACTGGGATAATGCACAGTTTTGGCTGACTTCAGTGACTGGGTCCAGTCC 1320
             ||||||||| |||||||||||||||||||| ||||||||||||||||||| ||||||||||
Db      4657 AGAGATGTTTGGGATTGGGAGATGCACGGTGGCTGTTGACTTCGGTGATTTCAGCCTCGTCC 4716

Qy      1321 AAGCTCCTGCGCGGATTACCGGGGAGTGCCCCTTTTTCTCAATGGCCAACGTGGGTACAAGGGG 1380
             ||||||||||||| |||||||||||||||| ||||||| |||| |||||||| ||||||||||
Db      4717 AAGCTTCCTGCGCGGGATTGCGGGGAGTCCCCTTTCTCATGGCCGAGGTGGGTACAAGGGA 4776

Qy      1381 GTCTGGCGGGAGACGGGCATCATGCAACACACCCTGCTCATGTGGACAACAGATCACCGGA 1440
             |||||||||||||| ||||||||| || |||||||| |||||||||| ||||||||||| ||
Db      4777 GTCTGGCGGCGGCATCATGCATGCACACCCCCATGTGGACGACAGAATCACCGGA 4836

Qy      1441 CATGTCAAAAAACGGTTCCATGAGGATCGTTGGGGCCTAAGACCTGTTAGTGAGTAACATGTGGCAT 1500
             |||||| |||||| |||||||||||| |||||||||||||| ||||||||||||| ||||||
Db      4837 CATGTAAAAAACGGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTCACGTGCAT 4896

Qy      1501 GGAACATTCCCCATCAAACGGATACACAGCGGGCCCCTGCCCTGCACCTCCCAGCGCCAAAC 1560
             ||||||||||||||| |||| ||||| ||||||||||| |||||||||||||||| ||||||
Db      4897 GGAACATTCCCCATTAACGGGTACACACGGGCCCCTGCCTGCACGGGCCCCTGCCCTGCACCTCCCAGCGCCAAAT 4956

Qy      1561 TATTCCAGGGCGGCGCTGGGTGCTGCTGCTGAGGAGTAGCCTGGGAGTTACGTGGGG 1620
             |||||||||||||| ||||||||||||||||| ||||||| |||| |||||||||
Db      4957 TATTCTAGGGCGTGTTGGGGCGGTGGCGTCTGTGAGGAGTACCTGGGAGTTACGGGGGG 5016

Qy      1621 GATTTCCACTACGTGACGAGCCATGACCACTGACAAGGTAAAATGCCCGTGCCGCCAGGTTCCA 1680
```

RESULT 8
AAA98967

```
ID   AAA98967 standard; DNA; 8001 BP.
XX
AC   AAA98967;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Hepatitis C virus DNA fragment SEQ ID NO: 3.
XX
KW   cell culture; therapy; infection; vaccine; diagnosis; gene therapy; ds.
XX
OS   Hepatitis C virus.
XX
PN   DE19915178-A1.
XX
PD   05-OCT-2000.
XX
PF   03-APR-1999;   99DE-01015178.
XX
PR   03-APR-1999;   99DE-01015178.
XX
PA   (UYMA-) UNIV MAINZ GUTENBERG JOHANNES.
XX
PI   Bartenschlager R;
XX
DR   WPI; 2000-629140/61.
XX
PT   Cell culture system for hepatitis C virus, useful e.g. in screening for
PT   therapeutic agents, comprises human hepatoma cells containing a viral RNA
PT   construct that includes a selectable gene.
XX
PS   Claim 7; Page 30-36; 58pp; German.
XX
CC   This invention describes a novel Hepatitis C virus (HCV) cell culture
CC   system comprising human hepatoma cells that contain an integrated HCV-RNA
CC   construct (I). (I) contains the HCV-specific RNA segments 5'-NTR (non-
CC   translated region), NS (non-structural)3, NS4A, NS4B, NS5A, NS5B and 3'-
CC   NTR, and a selectable (marker) gene (II). The cell cultures, and/or (I),
CC   are used to prepare, evaluate and/or test therapeutic and/or diagnostic
CC   agents for HCV infections, and to prepare vaccines against HCV infection
CC   (particularly preparation of attenuated HCV). The can also be used for
CC   preparation of a liver-specific delivery system for gene therapy, and to
CC   identify cells permissive for HCV replication. Virus RNA replicates
CC   autonomously and with high efficiency in this cellular system, so that
CC   variations in replication rates can be measured (for screening antiviral
CC   agents) quantitatively or qualitatively, using standard laboratory
CC   equipment. Efficient replication of HCV RNA is only achieved when the
CC   specified RNA segments are present and when the transfected cells are
CC   maintained under permanent selection pressure
XX
SQ   Sequence 8001 BP; 1652 A; 2373 C; 2244 G; 1732 T; 0 U; 0 Other;

Query Match             87.8%;  Score 3291.4;  DB 3;  Length 8001;
Best Local Similarity   92.4%;  Pred. No. 0;
Matches 3463;  Conservative    0;  Mismatches  286;  Indels  0;  Gaps  0;

Qy        1 TGGGAGGGGCGTCTTCACAGGCCTCACCACGTGGAATGCCCACTTCCTGTTCCCAAACAAAG 60
            |||||||| || ||||| |||| ||||||||  | |||||||||  |  ||||||| | ||||
Db     3409 TGGGAGGACGGTCTTTACAGGGCCTCACCACATAGACGCCCATTTCTTGTTCCCAGACTAAG 3468

Qy       61 CAGGCCAGGAGAGAACTTCCCCTACCTGGTGGCGTACCAGCCTACTGTGTGTCGCGTAGGGCC 120
            ||||| ||||||| ||||||||||||||||| || |||||| |||||| | ||| || |||||
Db     3469 CAGGCAGGAGACAACTTCCCCTACCTGGTAGCATACCAGGCTACTGTTCTTCGCGTCGGGGCT 3528

Qy      121 CAGGCCCCCACCCTCATCATTGGGATCAAATTGTGAAGTGTCTCATTACGGGTCACGGCCTACT 180
            |||||||||  ||  |||||   | ||||  |||| | |||||||| || |||||  ||||| |
Db     3529 CAGGCTCCACCTCACCTGTGGGACCAAAATGTGGAAGTGTTCATACGGGTCTAAAGCCTACG 3588

Qy      181 CTGGGCGGCGGCCAACACCCTTGCTGTATAGGCTGGGAGCCGTCCAAAAACGAGGTCACCCCTC 240
            |||| || ||||||  ||||||||  | ||| |  ||| |||||| || ||  ||| ||||| |
Db     3589 CTTCGACGCGGGCCAACGCCCTGCGTTATGGCTGGGATTGGATATGCACGGGTGTTGACCTGCC 3648

Qy      241 ACACACCCCATAACCAAAATTCATCATGGCATGCAGTCACCGGACCTGGAGGTCGTCGTCAGG 300
            ||||| |||||| | |||| | |||||||| || | ||||||| || |||| |  | |||||
Db     3649 ACACACCCCATAACCAAATTCATCATGGCATGCAGTCACCGGACCTGGAGGTCGTCGTCAGG 3708

Qy      301 AGCAACCTGGGTGCTCGGGTGGGGGCGGGGGTCCTTGCCAGCTCTGGCTGTATTGCTTGACAACA 360
            |||| |||||| ||| |||| ||||| ||||| || || |||| |||| | ||  ||||||| ||
Db     3709 AGCAACTTGGGTGCTCGGGTGCTCGGTAGCGGGAGTGTCCTAGCAGTCTTTGCCTGCGTGACAACA 3768

Qy      361 GGCACGGTGGTCATTGTGGGTAGGATCATCTTTGTCCGGGGGGGGGGCCCCGGCTTATTGTTCCGGAC 420
            |||| ||| || || |||||| ||||||||| || ||||| || ||||||| ||| ||||| |||||
Db     3769 GGCACCTGGGTCGCTGGTGGGAAGGATCATCTTGGGGTCGGTCCCGGGCCGGCATTTGCCTGACAACA 3828

Qy      421 AGGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGAGATGTCGCAGCACCTCCTTAC 480
            |||||||||| | ||||| ||||||||  ||||| || ||| | || |||  |||||  |||| |
Db     3829 AGGGGAAGTCCTTTACCAGGAGTTCGATGAGATGGAAGAGTGCGCATGTCCGGCACCTCCTTAC 3888

Qy      481 ATCGAGCAGGAAATGCAGCTCGCCAGGAGCTACGACGACCAGTTCAAGCAAAAAGCGTCGGGGTT 540
            |||||||||||| || |||| |||| ||||| | ||  |||||  |||| | |||||  | ||  |
Db     3889 ATCGAACACGGAAATGGCAGCTCGCCGAACAAATTCAAAACAGAAGGCAATCGGGTTGCAA 3948

Qy      541 ACAGCCACCAAGCAACGCGAGGAGGCCCGCTGCGTCCCGTGGTGGAGTCCAAGTGGCGAGCCCTT 600
            || |||||| ||||||| | || | || || | |||| ||| | |||| |||| ||||| |||||
Db     3949 ACAGCCACCAAGCAACGCGAGGAGGCTGCTCCGGTCAAAGTGGCGGAGCCCTC 4008

Qy      601 GAGGACCTTCTCGGGGCGGAAACACATGTGGGAACTTCATCAGCGGGGATACGAGTACTTAGCAGGC 660
            || |||||| | ||| |||||| |||| | | |||| ||||||||| | || | ||| |||||||
Db     4009 GAAGCCTTTCTCGGGGCGAAGCATTATGTGGGAATTTCAGCGGGGATACAAATATTTAGCAGGC 4068

Qy      661 TTGTCCACTCTGCCTGGGGAATCCCGGGGATTGCATCACACTGATAGCCGGTTCACAGCCCTGTC 720
            |||| |||||||| | |||||  |||| ||||| ||||| ||| |  || ||||| || |||||
Db     4069 TTGTCCACTCTGGCCTGGGAGTTCATCACTGATGACCGCGGGATTAGCATCAACTACTTCAGC 4128

Qy      721 ACTAGCCCGGCTCACCACCCAAATCTACCTCCTCTGCTTAACATCCTGCTTAGTTGGGATGGTAGCC 780
            || | |||||| |||| |||||  | ||| || | || ||| | ||||| | || | ||| ||||||
Db     4129 ACCAGCCCGGCTCACCACCCTGTTTAACATCCTGGTTGGGGATGGTAGCC 4188

Qy      781 GCCCCAACTCGGCTCCCCCCCAGTGGCTCGCTTCAGCTTTCGGTAGGGCCGTCATTGCTCGGTGCG 840
            |||| ||||||| ||| || ||||| | | ||||| |||||| |  ||| ||| |||| |  ||||
Db     4189 GCCCCAACTTGCTCCTCCCAGCCGCTCCTTCGGCTTTCGTAGGGCCCCCATGGGCATCGGCTGAGCG 4248

Qy      841 GCTGTTGGCAGCATAGGCCTTGGAAGGTGCTTGTTGGTGGACATCTTGGGCGGCGGGGGAGCA 900
            ||||||||||| || |||| |||| ||||| || |||| ||||| |  || ||| || |||||
Db     4249 GCTGTTGGCAGCATAGGCCTTGGAAGGTGTTGGTTGTGGAATTTTTGGCGGCGGGGGAGCA 4308

Qy      901 GGAGTGGCAGGCGGCCGCCTCGTGGGCCCTTTAAGGTCATGAGAGCGCGAGAAATGCCCTCCACGAG 960
            ||| || |||| || || | |||||| || ||||| |||||  | | |||| |||||  || |||||
Db     4309 GGGGTGGCAGGCGCGTCGGCCTTTAAGGTCATGAGCGCGCCTTCCACCGAG 4368

Qy      961 GACCTGGTTAACTTACTCCCTGCCATCCCTCTCCTGGTTGCCCTGGTGTCGTGTGTCCGGTG 1020
            |||||||| ||||| ||  || ||| |  || ||| || || |  ||||||| |||| ||||
Db     4369 GACCTGGTTAACCTACTCCCTGCATCCCTCTCCTGGCGCCCCTAGTGTGCGGGGGGTCGTG 4428

Qy     1021 TGGCGACGCGATACTCGGTCCGCACGTGGGTCCAGGGACGGCTGCAGTGGCAGTGGGAC 1080
            |||| ||| |||||| ||| |||| | || || ||| |||||  | ||||||| | |||
Db     4429 TGGCGACGGATAACTCGGTCCGGCACGTGGGCCCAGGGGACGGCCTGCAGTGGGAC 4488

Qy     1081 CGGCTGGATTAGCGGTTCGCCTTCCGCGGGGGTTACCACGTCCCACCGACTATGTGGCTGAGAG 1140
            |||| |||||| || ||| |||| || |||| |||||| || | ||| ||| |||||| || ||
Db     4489 CGGGCTGGATTAGGGTTCGGCCTTACCAGTCCCCACGGACGACTATGTGCCTGAG 4548

Qy     1141 AGCGAGCCCGGCAGCAGCGTGTCACTTAGATCCTCTCCGACCTTACTTATTCACCCAACTGTTTG 1200
            |||| |||| || |||| |   || |||||  ||| | ||| |||||| | ||||||| ||| ||
Db     4549 AGCGGACGCGCTGCGAGCGTGTCAGTGAACGTTACTCACACGACTATGTGCCTGCTG 4608

Qy     1201 AAGAGGCTTCCACCAGTGGACTTAACGGAAGGACTGGCTCCAAGCCTGCTGCGGCTCGTGTGGCTA 1260
            ||||| ||||| |||||| || |||| || | | | || ||||| || || | |||| ||| ||
Db     4609 AAGAGGCTTCACCAGTGGATCAACGGAAGGACTGCTCCAAGCCATGTTCCGGCTCGTGGCTA 4668

Qy     1261 AGGGGATGTTTTGGGACTGGAGATATGCACAGTTTTGGGTGACTTCAACGACTTCAGAGCCTCC 1320
            ||| |||| || ||||| | || |||||| | | || | |||| | ||| | || ||| ||||
Db     4669 AGAGATGTTTTGGGATTGGATAATGCACGGGTGTTGACTTCAAGACCTGGCTCCAGTCC 4728

Qy     1321 AAGCTCCTGCCGCGGGAGTTACCGGGAGTGCCGGGGGTACAAGGGG 1380
            ||| | || ||| ||| | ||| ||| || |  ||||| || ||||
Db     4729 AAGCTCCTGCCGCGGGATTGCCGGGAGTGGCCGGGGTAACAAGGGGA 4788
```

```
Db  6949  GACTTTGGTTCGTTATCTGTTGAAAGCGGGGGGGACCCAAGAGGAGAGGGGAGGGAGGCGGAGGCGGAGCCTACGGCC  7008
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3601  TTCACGGAGGCTATGACTAGGTACTCTCGCCCCCCCCCCCCCCCCCCCCAAACCAGAAATAC  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7009  TTCACGGAGGCTATGACTAGATAACTCCTGCCCCCCCCCCCCCAAACCAGAATAC  7068
          
Qy  3661  GACCTGCGAGTTGATGAATACATTCATGCTTCCTCCAAATGTGTCCGGTCCGGGCACCGATGCAATCATCTGGC  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7069  GACTTGGAGTTGATGAAACATCATGCTTCCTCCAAATGTGTGCGGTCCGCCGCACGATGCATCTGGC  7128
          
Qy  3721  AAAAGGGGTATATACCTCACCCGTGACCC  3749
          |||||||||||||||||||||||||||||
Db  7129  AAAAGGGGTGTACTATCTCACCCGTGACCC  7157
```

RESULT 9
ID AAA98966 standard; DNA; 8637 BP.
XX
AC AAA98966;
XX
DT 08-FEB-2001 (first entry)
XX
DE Hepatitis C virus DNA fragment SEQ ID NO: 2.
XX
KW Cell culture; therapy; infection; vaccine; diagnosis; gene therapy; ds.
XX
OS Hepatitis C virus.
XX
PN DE19915178-A1.
XX
PD 05-OCT-2000.
XX
PF 03-APR-1999; 99DE-01015178.
XX
PR 03-APR-1999; 99DE-01015178.
XX
PA (UYMA-) UNIV MAINZ GUTENBERG JOHANNES.
XX
PI Bartenschlager R;
XX
DR WPI; 2000-629140/61.
XX
PT Cell culture system for hepatitis C virus, useful e.g. in screening for
PT therapeutic agents, comprises human hepatoma cells containing a viral RNA
PT construct that includes a selectable gene.
XX
PS Claim 6; Page 23-29; 58pp; German.
XX
CC This invention describes a novel Hepatitis C virus (HCV) cell culture
CC system comprising human hepatoma cells that contain an integrated HCV-RNA
CC construct (I). (I) contains the HCV-specific RNA segments 5'-NTR (non-
CC translated region), NS (non-structural)3, NS4A, NS4B, NS5A, NS5B and 3'-
CC NTR, and a selectable (marker) gene (II). The cell cultures, and/or (I),
CC are used to prepare, evaluate, and/or test therapeutic and/or diagnostic
CC agents for HCV infections, and to prepare vaccines against HCV infection
CC (particularly preparation of attenuated HCV). The can also be used for
CC preparation of a liver-specific delivery system for gene therapy, and to
CC identify cells permissive for HCV replication. Virus RNA replicates
CC autonomously and with high efficiency in this cellular system, so that
CC variations in replication rates can be measured (for screening antiviral
CC agents) quantitatively or qualitatively, using standard laboratory
CC equipment. Efficient replication of HCV RNA is only achieved when the
CC specified RNA segments are present and when the transfected cells are
CC maintained under permanent selection pressure.
XX
SQ Sequence 8637 BP; 1768 A; 2558 C; 2436 G; 1875 T; 0 U; 0 Other;

Query Match
Best Local Similarity   87.8%;   Score 3291.4;
                92.4%;   Pred. No. 0;   DB 3;   Length 8637;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

```
Qy     1  TGGGGAGGGGCCTGTCTTCACGGACCTTCACCCAGTGGATGCCCACTTCCTTGTCCAAACAAAG  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4045  TGGGGAGGAGCGGTTCTTTACAGGCCCTCACCCACAATAGAGCCCCATTTCTTGTCCCAGACTAAG  4104

Qy    61  CAGGGCCAGGGAGGAGCAACTTCCCCTACGGTAGCCGTGGGGTAGCATACCAGGCTACGTGTGCGCCCAGGGCT  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4105  CAGGGCCAGGGAGCAGCAACTTCCCTACCTGGGTAGCATACCAGGCTACGTGTGCGCCCAGGGCT  4164

Qy   121  CAGGGCCCACCTCCATCTGTTGGGATCAAATGTTGGGAAGTGTTCATTAGGGCTAAAGCCTACT  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4165  CAGGGCTTCCACCTCCATCGTGGGACCAAAATGTGGAAGTGTTCATTAGGGCTAAAGCCTACG  4224

Qy   181  CTGCCGCGGCGGGCCAACACCCCTTTGCTTGTTGTATAGGCGTGGGAGGCCCTTCCAAAAACGAGGTCACCCTC  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4225  CTGCACGGGCCCAAACGGCCCTGCTGTTGTATAGGCGTGGGAGGCCCGTTCAAAAACGAGGTTACTACG  4284

Qy   241  ACACACCCCATAAACCAAATTCATCATCCATGGCCATGCAAGTTCAGCCCGACCTGGAGGTCGTCACG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4285  ACACACCCCATTAACCAAAATACATCATGGCAATGCCGTGACCCTGGAGGTCGTCACG  4344

Qy   301  AGCACCTGGGGTGCTGGTGTGGGGGGTCCTTGCAGCTTCTGGCGTTGGTGTATTGCTTGACAACA  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4345  AGCACCTGGGGTGCTGGTTAGGCGGGTGGAGTCCTAGCAGTCTTGGCGTTGGTGTATTGCTTGACAACA  4404

Qy   361  GGCAGCGTGGTCATTGTTGTGGGGTAGGATCACATCTTGTCCGGGGGCGGGGGCTTGGAGGGCTTGGCTGCGGAC  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4405  GGCACGGTGGTCATTGTGGGCAGGATCACATCTTGTCCGGGAAGCCGGGCCATCATTCCCGAC  4464

Qy   421  AGGGAAGTCCTCTACCAGGAGTTCGATGAGGATGAGAAGAGTGCGCGGTCGCACCTCCCTTAC  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4465  AGGGAAGTCCTTTACCGGGAGTTCGATGAGAATGAGAAGAGTGCGCGGTCGCACCTCCCTTAC  4524

Qy   481  ATCGGAGCAGGGAATCCAGCTCGCCCGAGCAGTTCAAGCAAAAAGCGCTCGGGTTGCTGCAA  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4525  ATCGAACAGGGAAATGCAGCTCGCCCGAACACAATTCAAACAGAAGGCAAATCGGGTTGCTGCAA  4584

Qy   541  ACACGCCACCAAAGCAAGCGAGAGGGCGCGCGTCGCTGCCCTGGGTGGGAGTCAAGGTGGGAGCCCTT  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4585  ACAGCCACCAAAGCAAGCGAGGGCTGCTCCCGGGTGAATCAAGTGGCCGGGACCCTC  4644

Qy   601  GAGACCTTCTGGGGGCGAAACACATGTGGGAACTTCATCAGGGGGGATACAGTGCTAGCAGGC  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4645  GAAGCCTTCTGGGGGGAAGCAATATGTGGGAATTTCATCAGCGGGATGGTACAAATATTTAGCAGGC  4704

Qy   661  TTGTCCACTCTGCCCTGGCAAATCCCCGGATTGCATCATCGATGAGCGGTTCAGGGCTCGTC  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4705  TTGTCCACTCTGCCCTGGCAACCCCGGATAGCACACTGATGGCATTCACAGGGCTCTATC  4764

Qy   721  ACTAGCCCCGCTCACACACCCAAATCTACCCTCTCCGCTTAAACATCCTTCGGGGGGGGATGGGTAGCC  780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4765  ACCAGCCCCGCTCACCCACCCAAACATACCCTCCTGTTTAACATCCTGGTTAACATCCGGGGGGATGGGTGGCC  4824

Qy   781  GCCCAAACTCGCCTCCCCCCCAGTGCTGCTTCAGCTTCGTTAGGCGGCCGCCGCAGGGGCTGCG  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4825  GCCCAAACTTGCTCCTCCAGCGCTGCTTCTGCTTTCGTTAGGCGGCCATCGCTGGCTGGAGCG  4884

Qy   841  GCTGTTGGCAGCATAGGGCCTTGGGAAGGTGCCTTGGGACATCTTCGGCGGCGCTATCGAGCA  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4885  GCTGTTGGCGAGCATAGGCCCTTGGGAAGGTGCCTTAAGGAGCCATGAGTGCAGTTATGGCAGCA  4944

Qy   901  GGAGTGCGCAGGGCGCGCCTCGTGGCCTTTAAGGTCATGAGGGGGCGAAATGCCCTCCACCGGAG  960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4945  GGGGTGGCAGGGCGCCCTCGTGGCTTTAAGGTCAATGAGCCGGAGATGCCTCCACCGGAG  5004

Qy   961  GACCTGGTTTAACTTAACTCCCCTGCCATCCTCTCCTGGTGTCCCCTGGTCGTCGGGGGTCGTG  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5005  GACCTGGTTTAACCTACTCCCCTGTATCTCTCCTCCTGGGGTCCCTTAGTTGCTCCGTCCGTG  5064

Qy  1021  TGCGCGAGCGGATACTCCGTCGGCCAGGTGGGGTTCCAGGGGACGGGAGGGGCGCGGGGCATGGATGAAC  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5065  TGCGCGGCGGATTACTGCGTCGGCGGCGCCCAGGGGGAGGGGAGGGGCTGTGTTGCAGTGGTGGATGAAC  5124

Qy  1081  CGGCTGATTAGGCGTTCGCCTCGCGGGGGTAACCATGTTTCCCCCAGCGACCACTATGTGCGCAGAG  1140
```

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which

DR P-PSDB; ABG30587.
DR WPI; 2002-575382/61.
PI Kukolj G, Pause A;
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
PR 22-DEC-2000; 2000US-0257857P.
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus APGK12-5'A replicon.
XX
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW cell culture replication; ds.
XX
OS Hepatitis C virus.
XX
FH Key   Location/Qualifiers
FH
FT mutation   replace(1,G)
FT   /*tag= a
FT CDS   1802..8407
FT   /*tag= b
FT   /product= "HCV NS2-5B"
FT   /note= "viral enzymes"
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This sequence represents hepatitis C virus replicon APGK12-
CC 5'A, created from the self-replicating HCV polynucleotide molecule APGK12
CC (see ABK88573)
XX
SQ Sequence 8638 BP; 1770 A; 2556 C; 2439 G; 1873 T; 0 U; 0 Other;

Query Match   87.8%; Score 3291.4; DB 6; Length 8638;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

Qy   1 TGGGAGGGGGCGTCTTCACAGGCCCTCACCCACGTGGTGATGCCCACTTCGTGTCCCAAACAAAG 60
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   4046 TGGGAGGAGCGTCTTTACAGGCCTCACCCACATAGACGCCCATTTCTTGTCTCCAGACTAAG 4105

Qy   61 CAGGCAGGAGGACAACTTCCCTACCTGGTAGAGTTCTCATACGGCTACCGTGTGCGCCCAGGGCT 120
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   4106 CAGGCAGGAGAGACAACTTCCCTACCTGGTAGCATACCAGGCTACGTGTGCGCCCAGGGCT 4165

Qy   121 CAGGCCCCACCTCCATCATGGGATCAAAATGTGGAAGTGTCTCATACGGGCTAAAGCCTACG 180
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4166 CAGGCTCCACCTTCCATCGTGGGACCAAAATGTGGAAGTGTTCTCATACGGGCTAAAGCCTACG 4225

Qy   181 CTGCGCGGCCGCCAAACACCCTTGCTGTATAGGCTGGAGAGCCTTCCAAAACGAGGTCACCTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4226 CTGCACGGGCGCCAAACGCCCCTCCTGTATAGGCTGGGAGAGCCGGTTCAAAAACGAGGTTACTACC 4285

Qy   241 ACACACCCCATAACCAAATTCATCATGGCATGCAATGTCAGGCGACGACCTGGAGGTCGTCAGG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4286 ACACACCCCATAACCAAATACATCATGGCATGCAATGTCAGGTCGCGGACCTGGAGGTCGTCACG 4345

Qy   301 AGCACCTGGGTGCTGGTGGGGGGGGTCCTTGCGGCTCGCGCTGCGTATTGCTTGACAACA 360
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4346 AGCACCTGGGTGCTGGTGGGGGGAGTCCTTGGCGGCGGTATTGCCTGACAAACA 4405

Qy   361 GGCCACGGTGGTCATTGTGGGTAGGATCATCTTGTCCGGGGCGCCGGCTATTGTTCCGGAC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4406 GGCCACGGTGGTCATTGTGGGCAGGATCATCTTGTCCGGGGCGCCGGCCATCATTCCGAC 4465

Qy   421 AGGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGAGTGCCGGCGTCGCACTTAC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4466 AGGGGAAGTCCTTACGGGGAGTTCGATGAGATGGAAGAGAGTGCGCGCCTTCACACCTTAC 4525

Qy   481 ATCGACGCAGGGAAATGCAGCTGCCCAGGCAGTTCAAGCAAAAAAGCGCTCGGGTTGCAG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4526 ATCGAACAGGGAAATGCAGCTGCCGAACAATTCAAACAGAAGGCAAAATCGGGGTTGCTGCAA 4585

Qy   541 ACAGCCACCAAGCAAGGCGAGGCCGCCGCTGCTCCGGTGGGTGAGTCCAAGTGGCGACCCTT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4586 ACAGCCACCAAGCAAGGCGAGGCTCGTGCTCCGGTGGAATCCAAGTGGCGACCCTC 4645

Qy   601 GAGAGCTTCTCGGGCGGAAACACATGTGGAACTTCATCAGGGGATAACAGTACTACAGAGGC 660
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4646 GAAGCCTTCTCGGGCGGAAGCATATATGTGGAATTTCATCACGCGGATACAAATATTTAGCAGGC 4705

Qy   661 TTGTCCACTCTGCCTGGAGAATCCCGCGGATTGGCATCACTGATGGGCGTTCACACTCCTGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4705 ...

RESULT 10
ID ABK88595   standard; DNA; 8638 BP.
XX
AC ABK88595;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus APGK12-5'A replicon.

Db   7765 AAAAGGGTGTACTATCTCACCCGTGACCC 7793

Qy   3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
          |||||||||||||||||||||||||||||
Db   7765 AAAAGGGTGTACTATCTCACCCGTGACCC 7793

Qy   3705 GACTTGGGAGTTGATAACATCATGTTCTCCACAATGTGTCAGTCACTCACTGGC 3720
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   7765 ...

Qy   3661 GACCTGGGAGTTGATAACATAGTGTCGTCGCCCACAGCTCATCCTGGC 3660
Db   7645 TTCACGGAGGCTATGGACTAGGTACTCCTGGTGCCCCCCCCCCACCCAGAATAC 7704

Qy   3601 TTCACGGAGGCTATGACTACTTCTGCCCCCCCCCCCGGGGGACCCCCCCCAAACCAGAATAC 3660
Db   7585 GACCTTGTCGTTATCTGTGAAAGCGCGGGGGACCCAAGAGGACGAGGCCTAGGGCGGCC 7644

Qy   3541 GGCCTTGTCGTTATCTGTGAGAGCGCGGGGAACCCAGGAGCAGCGGGGCCCGAGGCCTACGAGTC 3600
Db   7525 GCCCGCTGCCCTGTCGAGCTCGCCAGGACTTCTCGTATCGGGGGACT 7584

Qy   3481 GCCTCTGCAGCCCTGTCGTCGAAAGCTCCAGAGCTCGTCGTGTGTGCCGGAGAC 3540
Db   7465 TGCCGGCGCGACGGTGTTACTGACAGCTGTAATACCCTCACATGTTACTTGAAG 7524

Qy   3421 TGCCGCGCGAGCGCGTGTCGTGTGACCAGTACGTGTGCCGTPAATACCCTCACATTTCACTTGAAG 3480
Db   7405 CGGCCTTTACATCCGGGCGCCCCTGCAGAACTGCGGGGGCCATAAGGTCGGCCTATGCCCCG 7464

Qy   3361 CGGCGCGCCGAGCGACGCGTGTGTCGGGGGGGTCCCCTGACTAATTCAAAAGGGCAGAACTGCGGGGCTATCGCCGG 3420
Db   7345 ATCTACCAATGTTGTGACTTTGGTACTTGGCCCCCCGAAGCCAGACAGCCCATTAAGGTCGGCCTCACAGAG 7404

Qy   3301 ATTTATTCAATGTTGTTGTGACTTTGGCCCCCGGAAGCCCAGGCGGCCAATAAGGTCGGCCTCAGGAG 3360

Claim 11; Page 119-129; 140pp; English.

PS Ps
PT PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.

Db 4706 TTGTCCACTCTGCCTGGCAACCCGCGATAGCAATCACTGATGGCATTCACAGCCTCTATC 4765

Qy   721 ACTTAGCCCGCCTCACCACCAATCTTACCCTTCCTGCTTAACATCCTTGGGGGGGATAGCC 780
Db 4766 ACCAGCCCCGCCTCACCACCCAACATACCCTGTTTAACATCCTGTGGGGGGGATGGGTAGGGC 4825

Qy   781 GCCCAACTCGCTCTCCCCCCAGTCGTGCTTCAGCTTTCGTGTAGGGCGCGCCGGGCATTGCTGGTGCG 840
Db 4826 GCCCAAACTTGCTCCTCCCCAGCGCGTTGCTTCTTGCTTTCGTAGGGCGGGGCCCATCGGCTGGAGGG 4885

Qy   841 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTATGTGGGCATCTTGGCGGGCTATGGAGCA 900
Db 4886 GCTGTTTGGGCAGCATAGGGCCTTTGGGAAGGTGCTCTTGGCGATANTTTGGCAGGTTATGGGAGCA 4945

Qy   901 GGGAGTGGGCAGGGCGGGCGGCCTCGTGGGCCTTTAAGGTCATGAGCGGGGAAATGCCCTCCACCGAG 960
Db 4946 GGGGTGGGCAGGGCGGGCGGCCTCGTGGGCCTTTAAGGTCATGAGCGGGGGATGCCCTCACCGAG 5005

Qy   961 GACCTGGGTTAAACTTACTCCCGTGCCATCCTCTCTCCTTGCCATGAGCGGGGCACTATGGC 1020
Db 5006 CGGCTGATAGCGTTGCTTCGCGGGGGGTAACCACGTTCTCCCCACCGCACTATGTGCCTCGAG 5065

Qy 1021 TGCCCAGCGATAGGCGTCGGTCGGCACGTGGGTCCAGGCGGGGCGAAGTGCATGTTGATGGAAC 1080
Db 5066 TGCGGCAGCCGGATAGTGCGTTGCTTCGGGGGGTAACCGTGCCTCGGCACCGGATGGATGGAC 5125

Qy 1081 CGGCTGATAGGCGTTGCTTCGCGGGGGGTAACCACGTTCTCCCCACCGCACTATGTGCCTCGAG 1140
Db 5126 AGCCGACGCGCGGCAGCACGTGTCACTGGATAGTCCTCCTCCGACCTTACTATCACCCAACTGTTG 1185

Qy 1141 AGCCGACGCGCGGCAGCACGTGTCACTGGATAGTCCTCCTCCGACCTTACTATCACCCAACTGTTG 1200
Db 1261 AGGGATGTTTGGGGACTGGGATATGCACAGGTTTGGCGTGACTTCAAGACCTTGGCTCCAGTCC 1320

Qy 1261 AGGGATGTTTGGGGACTGGGATATGCACAGGTTTGGCGTGACTTCAAGACCTTGGCTCCAGTCC 1320
Db 5306 AGAGATGTTTGGGGATTGGGATATGCACGGTTGTTGACTGATTTCAAGACCTTGGCTCCAGTCC 5365

[The remainder of this page consists of a dense DNA sequence alignment (Qy query and Db database rows with flanking position numbers) that is not legibly resolvable.]

RESULT 11
ABK88573

ID   ABK88573 standard; DNA; 8639 BP.
XX
AC   ABK88573;
XX
DT   21-OCT-2002 (first entry)
XX
DE   Hepatitis C virus APGKI2 replicon.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key   Location/Qualifiers
FT   CDS   1803..8408
FT   /tag= a
FT   /product= "HCV NS2-5B"
FT   /note= "Viral enzymes"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G, Pause A;
XX
DR   WPI; 2002-575382/61.
DR   P-PSDB; ABG30581.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Example 1; Page 48-58; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations, transduces and/or replicates with greater
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This sequence represents hepatitis C virus replicon APGKI2, a
CC   self-replicating HCV polynucleotide molecule
XX
SQ   Sequence 8639 BP; 2556 C; 2441 G; 1873 T; 0 U; 0 Other;

Query Match   87.8%; Score 3291.4; DB 6; Length 8639;
Best Local Similarity   92.4%; Pred. No. 0; DB 6;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

Qy   1 TGGGGAGGGGCGTCTTCACAGCCCCTCACCCAGTGGGATGGCCACTTCCCA...AACAAAG 60
          ||||| || ||||||||| ||| |||||||||| ||| ||| |||| | ||||| ||
Db   4047 TGGGAGGACGCGTCTTTACAGGCCTCACCCACATAGACGGCCCATTTCTTGTCCAGACTAAG 4106

Qy   61 CAGGCAGGAGACAACTTCCCTACCTGGTGGCGGCCGTACCAGGCTACTGTGTGGGCGGGCC 120
          |||||||||| ||| ||||||||| |||||| || || ||||| ||||| | |||||||
Db   4107 CAGGCAGGAGCACAACTTCCCTACCTGGTAGCATACCAGGCTACCGGTGTGCGCCCACGGGCT 4166

Qy   121 CAGGGCCCACTCCATCATGGGATCAAAATGTTGGAAGTGTTCATACGGGCTAAAGCCTACG 180
          |||| || ||||| ||| ||| |||||||||||| ||||||||||| | ||||| | |
Db   4167 CAGGCTCCACCTCCATCGTGGGACCAAAATGTTGGAAGTGTTCTCATACGGGCTCACCCTC 4226

Qy   181 CTGCGCCGGGCCCAACCACCTTGTGTTATAGCGCGGTAGCCTGCCGCCCCAAAACGAGGTC 240
          || |||||||||||| || | || || |||| |||||||| | |||| ||||| ||| |
Db   4227 CTGCAGGGCGCCAACCGCCCTCCTGTATAGCCGGTGGGAGCCGGTTACTACC 4286

Qy   241 ACACACCCCATAACCAAATTCATCATGGCATGCATGTCAGCCGACGACTTGAGGTCGTCACG 300
          ||||| ||| || ||| || |||||||| ||| || || ||||||||| |||| | |||||
Db   4287 ACACACCCCATAACCAAATACATCATGGCATGCATGTCAGCCGACGACTTGAGGTCGTCACG 4346

Qy   301 AGCACCTGGGTGCTGGTGGGGGGCGGGTTCCTTGCAGCTCTTGGCGTATTGCTTGACAACA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ... AGCACCTGGGTGCTGGTGGGGGGCGGGTTCCTTGCAGCTCTTGGCGTATTGCTTGACAACA 360

[Sequence alignment blocks — Query (Qy) versus Database (Db), partial listing:]

Qy   2881 TCTAAATTTGGCTATGGGGGCAAAGGACTTCCGGAACCTATCCTCCGCAAGGCCCATTAACCAC 2940
Db   6926 TCTAAAATTTGGCTATGGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCCGTTAACCAC 6985

Qy   2941 ATCCGGCTCCGTGTGTGGGGGGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000
Db   6986 ATCCGGCTCCGTGTGTGGGGAAGGACTTGCTCTGAGAAGACACTGAACACCAGCCATC 7045

Qy   3001 ATGGCAAAAAATGAGGTTTTTCTGGGTCCAACCAGAGAGGAGGGCCGCAAGCCAGCTCGC 3060
Db   7046 ATGGCAAAAAATGAGGTTTTTCTGGGTCCAACCAGAGAGGAGGGCCGCAAGCCAGCTCGC 7105

Qy   3061 CTTATCCGTGTTCCCGACTTGGGGGTCCGGGTCTGCGAGAAAATGGCCCTCTATGACGTG 3120
Db   7106 CTTATCGTATTCCCGACTTTGGGGGTTCGTGTGCGAGAAAATGGCCCTTTACGATGTG 7165

Qy   3121 GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCAACCAGAGTATTCTCCCTGGA 3180
Db   7166 GTCTCCACCCTCCCTCAGGCCGTGATGGGCTCTTCATACGGGATTCCAATACTCTCCTGGA 7225

Qy   3181 CAGGGGGTCGAGTTCCTGGTGAACGACCCTGAAATCAAAAGACCCCTTCTATGACGTTGCA 3240
Db   7226 CACGGGGTCGAGTTCCTGGTGAATGCCTGGAAATGCCCTATGGGCTTTCGGCA 7285

Qy   3241 TATGACACCAGCCGTGTTTGACTCAAACAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300
Db   7286 TATGACACCCGCTCGTTTTGACTCAAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 7345

Qy   3301 ATTTATCAARTGTTGTGACTGGCCCCGAAGCCAGAGGCCATAAGGTCGCTCACAGAG 3360
Db   7346 ATCTACCAARTGTTTGACTTGGCCCCGAAGCCAGACGCCGCCATTAAGGTCGCTCACAGAG 7405

Qy   3361 CGGCTTTATATCGGGGTCCCCTGACTAATTCAAAAAGGGCCAGAACTGCGCGGCTATCGCCG 3420
Db   7406 CGGCTTTACATCGGGGCCCCCTGACTAATTCTAAAAGGGCCAGAACTGCTGCCCTATCGCCCG 7465

Qy   3421 TGGCGCGGGGAGCGCGGGGCGGTACGTGCGGTAGTAATACCCTCACATGTTTACTTGAAG 3480
Db   7466 TGCCGCGCCGAGCGGTGTACTGACCAGCGTGTAAATACCCTCACATGTTACTTGAAG 7525

Qy   3481 GCCTCTGCAGCCTGTCGAGCTGCAGAAAGCTTCCAGGACTGCAGATGCTCGTTGTGTGCGGAGAC 3540
Db   7526 GCCGGCTGCGGCCCGTGTCGAGCTGCGGAAAGCTTCAGATGCTCGTTGTGTGCGGAGAC 7585

Qy   3541 GGCCTTGTCGTTATCTGTTGTGAGACGGGCGGCGGGCGACGGCGGCGACGCCCTACGAGTC 3600
Db   7586 GACCTTGTCGTTATCGTTGTCGAAAAGCGGCGGGGACCAAGAGCGACGACCCTACGGGGCC 7645

Qy   3601 TTCACGCGGAGGCTATGGACTAGGTACTACTCTGCGCCCCGCCCCCCAAAACCAGAGAATAC 3660
Db   7646 TTCACGGGAGGCTATGACTAGGTACTACTCTGCCCCTGCCCCCGTCGGGGAACCAGAATAC 7705

Qy   3661 GACCTGGAGTTGATAACATCATGGTCTCCTCCAAATGTGTCGGTCGGGCCGCACAAGTGTCGGTG 3720
Db   7706 GACTTGGAGTTGATTAACAGTCATCAGTGTCCAGCCGCACGGATGCATCTGGC 7765

Qy   3721 AAAAAGGGTATACTACCTCACCCGTGACCC 3749
Db   7766 AAAAAGGGTGTACTACTACTCACCCGTGACCC 7794

```
Qy  2521  CCCGATTCTCAGCGACGAGGGGTCTTGGTCTTCTTACCGTGAGTGAGGAGGTCGTC  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6567  CCCGATTCTCAGCGACGAGGGGTCTTGGTCTACCGTGAAGCGAGGAGGTCGTC  6626

Qy  2581  TGCTTGCTTCGATGTCCTACACATGACACAGGCGGGCGGCTCTGATCACGCCATG  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6627  TGCTTGCTTCGATGTCCTACACATGACACAGGCGGGCCCTGATCACGCCGAGGAA  6686

Qy  2641  AGCAAGCTGCCATCAAACGGGTTGAGCAACTCTTTGCTCGTTCACCAACACATGGTCTAC  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6687  ACCAAGCTCGCCATCAAATGCACAGCGACTGAGCACTCTTTGCTCCGTCACCAACTTGGTCTAT  6746

Qy  2701  GCTACCACATCCCCAGCGCGCAAGCCCAGCGCGCAGAGAGAAGGTCACCTTTGACAGACTGCAA  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6747  GCTACAACATCCGCAGCGCGGCAAGCCTGCGCGCAGAAGAGGTCACCTTTGACGACTGCAG  6806

Qy  2761  ATCCTGGACGGATCACTACCAGGACACTGCTTCAAGGCGGGTCCACAGTT  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6807  GTCCTGGACGAGCACTACCGGGGACGTGCTCAAGGAGGAGATGAAGGTCGGAAGGGGCGTTCACAGTT  6866

Qy  2821  AAGGCTAAGCTTCTATCAGTAGAGAGGAAGCCTGCAAGGTCGAAGGTGCCAAA  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6867  ATCCGGCTCCGTGTGGGAGGACTTGTTGTTGGAAGACACTGAGACCACCACCATC  6926

Qy  2881  TCTAAATTTGGGCTATTGGGGGCGCAAAGGACCATTCCAGCAGGCCCATTAAACCAC  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6927  TCTAAATTTGGCTATGGGGCGCAAAGGACGGTCGGGACCTTACCAGCAAGGGCCGTTAACCAC  6986

Qy  2941  ATCCGCTCCGTGTGGGAGGACTTGTTGTTGGAAGACACTGAGACCACCACCATC  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6987  ATCCGGCTCCGTGTGGGAGGACTTCGTGTTGGAAGACACTGAGACCACCACCATC  7046

Qy  3001  ATGGCAAAAAATGAGGTTTTTCTGCGTCCAACCAGAGAGAGGAGGGGGCGCAAGCCCGC  3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7047  ATGGCAAAAAATGAGGTTTTTCTGCGTCCAACCAGAGAGAAGGGGGGCCGCAAGCCAGCTCGC  7106

Qy  3061  CTTATCGTGTTTCCAGACTTGGGGCGTCCCCTGTGTGTTGGCGCGAGAAGAAGCGTG  3120
          |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  7107  CTTATCGTATTCCAGATTTGGGGCGTCCTGTGTGTCGGCGAGAAGAATGGCCCTTTACCAATGTG  7166

Qy  3121  GTCTCCACCCTCCCTCAGGCTGTGANTGGGCCTCTCGTACGGATTCCAGTATATTCTCCTGGA  3180
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  7167  GTCTCCACCCTCCCTCAGGCCGTGATGGGGCTCTTCATACGGATTCCAATACTCTCCTGGA  7226

Qy  3181  CAGGGGGTCGAGTTCCTGGTGAACCCTGAAATCAAAAGAGAACCCCTATGGGGCTTTGCA  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  7227  CAGGGGGTCGAGTTCCTGGTGAATGCCCTGGAAAAGCGAAGAAATGCCCCTATGGGGCTTCGCA  7286

Qy  3241  TATGACACCGGTGTTTTGACTCAACAGTCACTCAGAGATGACATCGGTAGAGGAGTCA  3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7287  TATGACACCGGCGTTTTGACTCACTGAATCACATCCGGTGTTGAGGAGTCA  7346

Qy  3301  ATTTATCAATGTTGTGACTTGGACCCGAAGCCAGAGCCCATAAGGTCGCTCTACAGAG  3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7347  ATCTACCAATGTGTGACTTGGACCGGAAGCCAGACCAGGGCCCATATAAGGTCGCTCACAGAG  7406

Qy  3361  CGGCTTTATATCGGGGTCCCCTGACTAATTCAAAAGGGGCAGAACTGCGGGTATCGCCGG  3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7407  CGGCTTTACATCGGGGGCCCCCTGACTAATTCTAAAAGGGGCAGAACTGGCCTATCGCCGG  7466

Qy  3421  TGCCGCGGCGGAGCGGCGTGCTGACGGCGTAATACCCTTCACATAGTTACTTGAAG  3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7467  TGCCGCGGCGGAGCGGGCGGCGGTGTTGACTGACCAGCTGCGTAATCCCTTCACATTACTTGAAG  7526

Qy  3481  GCCTTCTGCAGCCTGTCGAGCTGCAAAGCTTCCAGGACTGCGACTGCACGATGCTCGAGC  3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7527  GCGCGCTGCGGCCGTGTCGGAGCTGCGAAAGCTTCCAGGACTGCACGACTGCTCTCTATGCGGAGAC  7586

Qy  3541  GGCCTTGTTGTGTTATCTGTTGAGAGCGGGCGGAACCCAGGAGGACGCCTTACGAGTC  3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7527  GCGCGCTGCGGCCGTGTCGGAGCTGCGAAAGCTTCCAGGACTGCACGACTGCTCTCTATGCGGAGAC  7586

Qy  3601  TTCACGGAGGCTATGACTAGGTACTCTGCCCCCCGGGGGACCCGCCCCAACCAGAATAC  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7587  GACTTGGTCTGTTATCTGTGAAAAGGCGGGGGACCACAAGAGAGACGAGGGGCC  7646
```

RESULT 12
AAA98969
ID   AAA98969 standard; DNA; 8649 BP.
XX
AC   AAA98969;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Hepatitis C virus DNA fragment SEQ ID NO: 5.
XX
KW   Cell culture; therapy; infection; vaccine; diagnosis; gene therapy; ds.
OS   Hepatitis C virus.
XX
PN   DE19915178-A1.
XX
PD   05-OCT-2000.
XX
PF   03-APR-1999; 99DE-01015178.
PR   03-APR-1999; 99DE-01015178.
XX
PA   (UYMA-) UNIV MAINZ GUTENBERG JOHANNES.
XX
PI   Bartenschlager R;
XX
DR   WPI; 2000-629140/61.
XX
PT   Cell culture system for hepatitis C virus, useful e.g. in screening for
PT   therapeutic agents, comprises human hepatoma cells containing a viral RNA
PT   construct that includes a selectable gene.
XX
PS   Claim 9; Page 44-50; 58pp; German.
XX
CC   This invention describes a novel Hepatitis C virus (HCV) cell culture
CC   system comprising human hepatoma cells that contain an integrated HCV-RNA
CC   construct (II). (II) contains the HCV-specific RNA segments 5'-NTR (non-
CC   translated region), NS (non-structural)3, NS4A, NS4B, NS5A, NS5B and 3'-
CC   NTR, and a selectable (marker) gene (II). The cell cultures, and/or (I),
CC   are used to prepare, evaluate and/or test therapeutic and/or diagnostic
CC   agents for HCV infections, and to prepare vaccines against HCV infection
CC   (particularly preparation of attenuated HCV). The can also be used for
CC   preparation of a liver-specific delivery system for gene therapy, and to
CC   identify cells permissive for HCV replication. Virus RNA replicates
CC   autonomously and with high efficiency in this cellular system, so that
CC   variations in replication rates can be measured (for screening antiviral
CC   agents) quantitatively or qualitatively, using standard laboratory
CC   equipment. Efficient replication of HCV RNA is only achieved when the
CC   specified RNA segments are present and when the transfected cells are
CC   maintained under permanent selection pressure
XX
SQ   Sequence 8649 BP; 1773 A; 2563 C; 2437 G; 1876 T; 0 U; 0 Other;

Query Match   87.8%; Score 3291.4; DB 3; Length 8649;
Best Local Similarity   92.4%; Pred. No. 0;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

```
Qy     1  TGGGAGGGGGCGTCTTTCACAGGCCCTCACCCACGTGAATGCCCACTTCCTGTCCCAAACAAAG    60
          |||||||||||||||||||||||||||||||||||||||||||
Db  4057  TGGGAGGAGGCGGTCTTTACAGGGCCTCACCCACACATAGAGCGCCCATTTCTTGTCCCAGACTAAG  4116
```

```
Db  6277  TTAGAGTCCTGGAAGGACCCGGGACTACGTCCCTCCAGTGGTACACGGGGTGTCCAGTGCCG
Qy  2281  CCTACTAAGACCCTCCTATACTACCACCTCCACGGAGGAAGAGCAGTTGTTCTGACAGAA  2340
          ||||| |||| |||||||| |||||||||||||||||||||||||||| | ||||||||||
Db  6337  CCTGCCAAGGACCCCTCCGGATACCACCTCCACGGAGGAAGAGGACACGTTGTCCTGTCAGAA  6396

Qy  2341  TCCACGGTGTCTTCTGCCCTGGCCGGAGCTTGCCACAAAGGCTTTCGGTTAGCTCCGAACCG  2400
          ||||||||||||||||||||||||
Db  6397  TCTACCGGTGTCTTCTGCCCTTGGCCGGAGCTGGCCACAAAGACCTTGGGGCAGCTCCGAAATCG  6456

Qy  2401  TCCGCCCGTTCGACACCGGACACCGGAGCCAACCTGCTCTCCGACCACCGGCGGA  2460
          |||||||
Db  6457  TCCGGCCCGTCGACACGGGCACGCGGAAGAACCCCTCCTGACCACCGGCGGGGAC  6516

Qy  2461  GCAGGGATCTGACGTTGAGTGTATTCCTCCACCTCCAATGGCCCGCCGAGTC  2520
          |||||||
Db  6517  GCCGGGATCCGGACGTTGAGTCGTACTCCTCCATGCCCCCCTTGAGGGGGGGGAT  2576

Qy  2521  CCCGATCTCAGCGAGCGGGTCTTGTTCTTACCGTGAGTGAGGGAGGTCGTC  2580
          |||||||
Db  2577  CCCGATCTCAGCGACGGGGTCTTGGTCTACCGGTAAGCGAGGGAGGACGTCGTC  2636

Qy  2581  TGCTGGCTCGGATGTGTCCTACACAGGACAGGGGGCTCTGGAGGAA  2640
          |||||||
Db  2637  TGGCTGCTCGGATGTCCTACACATGGACGGGGGCCTGGGAGGAA  2696

Qy  2641  AGCAAGGTCTCCCATCAAACGCGTTGAGCAACTCTTTTGCTCGGTCTAC  2700
          |||||||
Db  2697  ACCAAGCTCCCCATCAAATGCACTGAGCAACTCTTTTGCTCCGTCTAT  2756

Qy  2701  GCTACCACCATCCCGCGCAGGGCCAGCGCGACGAGAAGAAGGTCACTTTTGACAGACTGCAA  2760
          |||||||
Db  2757  GCTACAACATCTCCGCAGGGGCAAGCCTGGGCAGAAGAAGGTCACTTTTGACAGACTGCAG  2816

Qy  2761  ATCCTGGACGGATCACTACCAGACGGGACGGCGTCTCAAGGAGAGATCCGGCCAA  2820
          |||||||
Db  2817  GTCCTGGACGGACCACTACCGGGACGTGCTCAAGGAGAGAGGCGTCACAGTT  2876

Qy  2821  AAGGCTAAGCTTCTATCAGTAGAGAAGGAAGCCTGCAAGGAGAGACTCGGCGCAA  2880
          |||||||
Db  2877  AAGGCTAAACTTCTATCCCTGGAGGAAGCCTGTAAGTCTGAGCGCCCCACACATTCGGCCAGA  2936

Qy  2881  TCTAAAATTTGGGCTATGGGGCAAAGGACGTCCGGCAAACCTATCCAGCAGGCCCATTAACCAC  2940
          |||||||
Db  2937  ATGGCAAAAAAATGAGGTTTTCTGCGTCAACCAGAGAAGGGGCCGCGAAAGCCAGCTCGC  2996

Qy  2941  ATCCGCTCGGGGAGGACTTGTTGGAAGACACTCCAAATTGACACCACCATC  3000
          |||||||
Db  2997  ATCCGCTCGTGTGGAAGGACATTGCTGGAAGAGACACTGAGACACCACCATG  3056

Qy  3001  ATGGCAAAAAATGAGGTTTCTGCGTCCAACCAGAGAAGGCGCCGCAAGCCCAGCTCGC  3060
          |||||||
Db  3057  ATGGCAAAAAAATGAGGTTTTCTGCGTCAACCAGAGAAGGGGCCGCGAAAGCCAGCTCGC  3116

Qy  3061  CTTATCGTGTTCCCAGACTGTTGGGGTCCGGTCCGTGGAAAATGGCCCTCTCATTGACCTG  3120
          |||||||
Db  3117  ATGGCAAAAAAATGAGGGTTTTCTCGCGTCAACCAGAGAAGGCGCCGCAAGCCCAGCTCGC  3176

Qy  3121  GTCTCCACCTCCCCTCAGGCCTGTCGATGGGGTCAGTGTACTGCTTACTCTCCTGGA  3180
          |||||||
Db  3177  GTCTCCACCTCCCTCAGGCCGTGATGGGGTCTCTTCATCAGTGTATTCTTCCTGGA  3236

Qy  3181  CACGGGGTCGAGTTCCTGGTGAACGCCTCGAAATCAAAGAGAGACCCTAGGCTTTGCA  3240
          |||||||
Db  3237  CAGGGGGTCGGAGTTCTGGTGAACTGCGGAAAGCGAACGACAGACCCTATGGGGGCTTGCA  3296

Qy  3241  TATGACACCCGCCTGTTTTTGACTTGGCCCCCGAAGCCAGAGACAGGCCCATAGTTCCGGTGAGGAGTCA  3300
          |||||||
Db  3297  CAGCGGGTCGGAGTTGCTGGGATGAATGCCCGAACGACAGAAATGCCTATGGGGCTTCGCA  3356

Qy  3301  ATTTATCAATGTTGTGTGACTTGGGCCCCCGAAGCCAGAGACAGGCCCATAGTTCGCTCACAGAG  3360
          |||||||
Db  3357  TATGACACCCGCCTGTTTTGACTTGGCCCCGAAGCCAGAGACAGGCCCATAAGGTTCGCTCACAGAG  3416
```

RESULT 13
ABK91424

```
ID   ABK91424 standard; DNA; 9605 BP.
XX
AC   ABK91424;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA mutant 1.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication; mutant.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             342..9374
FT                   /tag= a
FT                   /product= "HCV polyprotein"
FT                   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT                   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
FT   mutation        .
FT                   replace(3625,G)
FT                   /tag= b
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
```

PT expression.
XX
PS Claim 9; Page; 69pp; English.
XX
CC The invention relates to nucleic acid molecules comprising altered HCV
CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC internal ribosome entry site (IRES) region coding for one or more NS3,
CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC are detailed in the specification. Also included are (1) an expression
CC vector comprising a nucleotide sequence coding for the altered nucleic
CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC recombinant cell human hepatoma cell comprising the altered nucleic acids
CC ; (3) a recombinant cell produced by introducing into a human hepatoma
CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC replicon enhanced cell or which containing a functional HCV replicon; (5)
CC an HCV replicon enhanced cells made in the method; and (6) measuring the
CC ability of a compound to affect HCV activity. The HCV replicons and HCV
CC replicon enhanced cells are useful in studying HCV replication and HCV
CC replication, and HCV and host cell interactions, producing HCV RNA and
CC proteins, and providing a system for measuring the ability of a compound
CC to modulate one or more HCV activities e.g. to discover drugs which may
CC treat HCV mediated diseases such as liver failure, cirrhosis and
CC hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
CC mutant of the invention. Note: The present sequence is not shown in the
CC specification but was created by the indexer using the HCV sequence
CC appearing as ABK91411 and the information in Claim 9
XX
SQ Sequence 9605 BP; 1910 A; 2884 C; 2732 G; 2079 T; 0 U; 0 Other;

Query Match   87.8%; Score 3291.4; DB 6; Length 9605;
Best Local Similarity   92.4%; Pred. No. 0;
Matches 3463; Conservative   0; Mismatches 286; Indels   0; Gaps   0;

Qy   1 TGGGAGGGGGGTCTTCACAGACCCTCACCCACGTGGATGCCCACTTCCTGTCCCAACAAAG 60
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   5013 TGGGAGGAGCGTCTTTTACAGGCCCTCACCCACACATAGACGCCCATTTCTTGTCCCAGACTAAG 5072

Qy   61 CAGGCAGGAGGAGACAACTTCCCCTACCTGGTGCGTACCAGAGCACCCATTTCAGGCCTAGGGCC 120
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5073 CAGGCAGGAGGAGACAACTTCCCCTACCTGGTGCGTACCAGGCTACTGGTGTCGCCCAGGGGCT 5132

Qy   121 CAGGCCCCACCTCCATCATGGGATCAAAATGTGGAAGTGTCTCATATACGGGCTAAAGCCTACT 180
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   5133 CAGGCTCCACCTCCATCGTGGGACCAAAAATGTGGAAGTGTCCATACGGGCTAAAGCCTACG 5192

Qy   181 CTGCCGCGGGGCCCAACACCCTTGCTGTATAGGCTGGGAGCCGTTCAAAAACGAGGTCACCCTC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5193 CTGCACGGGGCCCAACACCCCTGCTGTATAGGCTGGGAGCCGTTCAAAAACGAGGTTACTACC 5252

Qy   241 ACACACCCCCATAAACCAAAATTCATCATGGCAATGTCAGCCGACCTCGAGGTGCTCACG 300
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   5253 ACACACCCCATAAACCAAAATACATCATGGGCATGTCGGCCGACGTGCTCACG 5312

Qy   301 AGCACCTGGGGTGCGTGCTTGGGTGGGGGGTCCTTGCAGCTCTGCGTATTGCTTGACAACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5313 AGCACCTGGGGTGCGTGGCTGGGGGACTCCTTAGCAGTCCTGCCGGCTATTGCCCTGACAACA 5372

Qy   361 GGCACGCGTGGTCATTGTGGGTTAGGGATCATCTTGTTGTCCCGGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||
Db   5373 GGCACGGTCGTGTCATTGTGGGCCAGGATCATCTTGTTCCCCGAC 5432

Qy   421 AGGGAAGTTCCTCTACCAGAGATTGCATGGATGACAGAGTGGCGCGGGTTCCGGCTCGGCACCTCCCTTAC 480
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   5433 AGGGAAGTTCTTTACGGGAGTTCGATGACGTGCGCGGTTCGCGGAGTGTGCTTCACCTCCCTTAC 5492

Qy   481 ATCGAGCAGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAGCGCCTCGGGGTTGCTGCAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5493 ATCGAACAGGGAAATGCAGCTCGCCCGAACAAATTCAAACAGGAGGCAATCGGGTTCGGGGTGG 5552

Qy   541 ACAGCCACCAAGCAACGCGAGGGCCCGGGTTCAAGTGTGGCGAGCCCTT 600
              |||||||||||||||||||||||||||||||||||||||||||
Db   5553 ACAGCCACCAAGCAACGGGAGGGCTGCTCCCGTGGTGGAATTCAAGTGGCGGGGACCCTC 5612

Qy   601 GAGACCTTCTGGGCGGAAAAACATGTGGAACTTCATCAGGCCTTCGTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db   5613 GAAGCCTTCTGCGGCGGAAGCATATGTGGGGAATTCATCAGGCCTTCAGAGCTCTGTC 5672

Qy   661 TTGTCCACTCTGCCTGGGGAATCCCGGATTGCATCAGTCATGGGCGCTCTATC 720
              ||||||||||||||||||||||||||||||||||||||||||
Db   5673 TTGTCCACTCTGCCTGGCAACCCGCGGATAGCATCACACTGCATGGGCATTCACAGCCTCTATC 5732

Qy   721 ACTAGCCCCGGTCACCACACCCAAATCTACCCTCGCTTAACATCCTGCGTTAACATCCTCTTGG 780
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   5733 ACCAGCCGGTCACCACCCAACATACCCTCGTTAACATCCTGTGGGGGGGGATGGGTGGCC 5792

Qy   781 GCCCAACTCGGCTCCCCCCAGTGCTGCTTCAGCGTTGGTAGGCCGCATTGTGCTGGCG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5793 GCCCAAACTTGCCTCCCCGAGGTGCTTCGGCTTGGTGTGAGCTTGGCTATGGAGCA 5852

Qy   841 GCTGTTGGCAGCATAGGGCCTTGGGAAGGTGCTTGTTGGTGGATATTTGGCAGGTTATGGAGCA 900
              |||||||||||||||||||||||||||||||||||||||||||||||
Db   5853 GCTGTTGGCGAGCATAGGCCCTTGGGAAGGTGCTTGTTGGTTGCGAGTATTTGGCAGGTTATGGAGCA 5912

Qy   901 GGAGTGGCAGGGCGGCGGCCTCGTGGGCCCTTTAAGGTCATGAGGGCGAAATGCCCTCCACGGAG 960
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   5913 GGGGTGGCCAGGGGGCGGCCTCGTGTGGCCTTTACGTCATGAGCGCGAGATGCCCTCCACGGAG 5972

Qy   961 GACCTGGTTTAACTTACTCCCGTGCCATCCTCTCTCCGTGGTTCCGCTGGTGTCGTCGGGGTCGTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5973 GACCTGGTTTAACCTACTCCCCTGCCATCCTCTCTCCGTGGTTCCGCTGGTGTCGTCGGGGTCGTG 6032

Qy   1021 TGCGCGAGCGATACTTCGGTCGGCACCTGGGGTCCAGGGGACGGGGGCTGCAGTGGATGGAAC 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6033 TGCGCGAGGAATACTTCGGTCGGCACGGTGGGGGCCCAGGGGAGGGGGCTGTGCAGGCGGATGGAC 6092

Qy   1081 CGGCTGATAGGGTTGCTTCGCCTCGCGCGGGGGGTAACGTAGAACGTCTCCCCACCGACTATGTGCTGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6093 CGGCTGAATACGGTTGCTTCCCGCCCGGGTTAACCGCACGTCTCCCCACCGACCACTAGTGGCTGAG 6152

Qy   1141 AGCGACGGCCCGCAGACGTCACTCAGATCCTCTCCGAGCCTTACTATCACCCAAACTGTTG 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6153 AGCGACGGCTGCAGCACGTGTCACTCGGATCCTCTCTAGTCTTTACCATCACTCAGCTGCTG 6212

Qy   1201 AAGAGGGCTCCACCAGTGGATTAACGGAGACTCCTCCAGCGCCTGCTCCGCTGCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6213 AAGAGGCTTCCACCAGTGGATCAAACGGAGACTGCTCCCGCCATGGCTCCGCTCGTGGCTA 6272

Qy   1261 AGGGATGGTTTGGGGACTGGATAATGCACAGTTTGGCTGACTTCAGAACCTGGGCTTCCAGTCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6273 AGAGAGTGTTTGGGATTGGGAATGCACGGGTGTTGACTTGATTTCAGACCTGGTCTTCCAGTCC 6332

Qy   1321 AAGCTCCTGCCGCGGATTACCGGGGAGTCCCCTTTTTCTCATGCCAACGTGGGGTACAAGGGGA 1380
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   6333 AAGCTCCTGCCGGCGGAATTGCGGGGAGTCCCCCTTCTTCTCATGTGAACGTGGGGTACAAGGGA 6392

Qy   1381 GTCTCGGCGGGAGAGAGGCCATCATGCAAAACACCCGCCATGTGGAGCACAGGATCACCGGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6393 GTCTGCGCGGGGCGAGGGCGATCATGCGAAAAAACACCCCATGTGGAGCACAGGATCACCGGA 6452

Qy   1441 CATGTCAAAAACGGTTCCATGAGGAATCGTTGGGGCCTAGGATCCTTAGTAAGCATGTGGCAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6453 CATGTGAAAAACGGTTCCATGAGGATCGTTGGGGCCTAGGAGTACACGTGGGCAT 6512

Qy   1501 GGAACATTCCCCATCAAACGCATACACCACCGGGCCCCTGACGCGGCCCAAAC 1560
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   6513 GGAACAATTCCCCATTAAAGCGATACACCACGGGGCACGTTATACGTGGCCCAAAT 6572

Qy   1561 TATTCCAGGCGCGCCTGTGTCGCGTGGCGTGCTCTGCAGGAGTTACGTGGAGGTTGGGGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6573 TATTCTAGGGCCGCTCTGTGGCGGTGGCCTGTGGAGTTAGCGGGTTGGGGG 6632

Qy   1621 GATTTCCACTACGTGACGAGGCAGCATGACAGTGACACTGACCAAGGTAAAAATGCCGGGTTCCA 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6633 GATTTCCACTACGTGACGCGGCCAATGACCAGTGACACTGACCAAGGTAAAAGTGCCGGGTTCAGGTTCCG 6692

Qy   1681 GCCCCCGAATTCTTCACGAAGTGGATGGGGGTGCGGCCGCTGCACAGGTTGGACGTTCCGGCGCGTGC 1740

RESULT 14
ABK91411
ID   ABK91411 standard; DNA; 9605 BP.
XX
AC   ABK91411;
XX

DT   15-NOV-2002 (first entry)
XX
DE   Hepatitis C virus Con 1 isolate DNA.
XX
KW   HCV; ds; Con 1; adaptive mutation; liver failure; cirrhosis;
KW   hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW   internal ribosome entry site; IRES; NS5A; HCV replication.
XX
OS   Hepatitis C virus.
XX
FH   Key   Location/Qualifiers
FT   CDS   342..9374
FT   /*tag= a
FT   /product= "HCV polyprotein"
FT   /note= "The polyprotein consists of the Core, E1, E2, P7,
FT   NS2, NS3, NS4A, NS4B, NS5A and NS5B proteins"
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 9; Page 36-39; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is the HCV replicon Con 1,
CC   used as a basis for the adaptive mutations of the invention
XX
SQ   Sequence 9605 BP; 1910 A; 2883 C; 2733 G; 2079 T; 0 U; 0 Other;

Query Match   87.8%;   Score 3291.4;   DB 6;   Length 9605;
Best Local Similarity   92.4%;   Pred. No. 0;
Matches 3463;   Conservative   0;   Mismatches 286;   Indels   0;   Gaps   0;

Qy   1 TGGGAGAGGGGCGTCTTCACAGAGCCTCACCCACCACGGTGATGAATGCCCACTTCCTTGTCTCCCAAACAAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5013 TGGGAGGAGCCTCTTTACAGGGCCTCACCCACATAGACGCCCACTTTCTTGTCTCCAGACTAAG 5072

Qy   61 CAGGCAGGAGGAGACAACTTCCCTACCTGGTGGTCGGTAGCATACCAGGCTACGGTGGTGCGGCTAGGGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5073 CAGGCAGGAGGAGACAACTTCCCTACCTGGTGGTCGGTAGCATACCAGGCTACGGTGGTGCGGCTAGGGCT 5132

Qy   121 CAGGGCCCACCTCCACTCCATCATGGGATCAAAATGTGGAAGTGTCTCATACGGCTACTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5133 CAGGGCTTCCACCTTCACCTCCATCGTGGGACCAAAATGTGGAAGTGTCTCATACGGCTACG 5192

Qy   181 CTGCGGCGGCCAACACCCTTGCTTGTATAGGCGTCGGACGCCCTCAAAAACGAGGTGCACCTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5193 CTGCACGGGCGCCAAGCGCCCTTGCTTGTATAGGCTCGGACGCCGTTCAAAAACGAGGTTACTACC 5252

Qy   241 ACACACCCATAACCAAAATTCATACATGGGATCAATGGGAAGTGTGGAAGTGTCCTCACG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5253 ACACACCCCATAACCAAAATACATCATGGGAATGCATGTCACGCCGCTGACCCTGGAGGTGCTCACG 5312

Qy   301 AGCACCTGGGTGCTGGTGGTGGGCGGGGGTCCTTGCAGCTTCTGGCTGCTGTATTCCTTGACAACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5313 AGCACCTGGGTGCTGGTGGTAGGCGGGAGTCCTTAGCAGCTTCTGGCTGCTGTATTTGCCTTGACAACA 5372

Qy   361 GGCACGGTGGTCATTGTGGTAGGATCATCATTGTCGTCCGGCGGCGGGGCGGGGCTATTGTTCCCGGAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5373 GGCACGGTGGTCATTGTGGCCAGGAGCATCACTTGTCCGGAAAGCCCGGCCATCATTCCCGAC 5432

Qy   421 AGGGAGAAGTCCTTCTACCAGGAGTTCGATGAGATGGAAGAGTGCGGCTCGCAGCCCTTAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5433 AGGGAGAAGTCCTTTACCGGGAGTTCGATGAGATGGAAGAGTGCGGCTCACACCTCCCTTAC 5492

Qy   481 ATCGAGCACGGGAATGCAGCTCGCCCAGCACTTCAAGCACAAAAAGCGCTCGGGTTGCTGCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5493 ATCGAACACGGGAAATGCAGCTTCGCCCGAACAATTCAAACACAGAAGCAATCGGGGTTGCTGCAA 5552

Qy   541 ACAGCACCAAGCAAGCAGGAGGGCCGCTGCTGTCCCCGTGGGTGAGTCCCCAAGTGGGCGCCCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5553 ACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCCGTGGGTGAGTCCCCAAGTGGGCGACCCTC 5612

Qy   601 GAGACCCTTCTGGGCGAAAACACATGTGGAACTTCATCAGCGGGGATACAGTACTTAGCAGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5613 GAAGCCTTCTTCTGGGGGAAGCAATATGTGGAAATTTCATCAGCGGGGATACAAATATTTAGGAGGC 5672

Qy   661 TTGTCCACTCTGCCTGGGGAATCCGCGCGATTGCATCACTGATGGCGTTCACACCTGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5673 TTGTCCACTCTGCCTGGGCAACCCCGCGATTAGCATCACTGATGGCGTTCACAGCCCTCTATC 5732

Qy   721 ACTAGGCCGGCTCACCAACCCARTCTCACCCTCCTCTGTTAACATCCTGAGAGCAGTGACAGTATC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5733 ACCAGCGCCGTCACCAACCATACCCTCCTGTTAACATCCTGGGAGTGAGATGGGTTGGCC 5792

Qy   781 GCCCAAACTCGCTCCCCCAGTGGCTGCTTCAGCTTTCGTAGGCGTAGGCGGCCGCATTGCTGGTGGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5793 GCCCAACTTGCTCCTCCAGCGCTTCTTCGTAGGCGTAGGCGGCCGCATCGCTGGCGCG 5852

Qy   841 GCTGGTTGGGCAGCAATAGGGCCTTGGGAAGGTGCTTGGTGTGGACAATCTTGGCGGCGTATGGAGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5853 GCTGGTTGGGCAGCATAGGGCCTTGGGAAGGTGCTTGGTGTGGATAATTTTGGCAGTTATGGAGCA 5912

Qy   901 GGAGTGGCAGGCGCCGCTCGTGGTGGCCCTTAAGGTCATGAGCGGGCGAAAATGCCCTCCACGGAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5913 GGGGTGGCAGGCGCCGCTCGTGGTGGCCCTCAAGGTCATGAGCGGGGAGAGTGCCCTCCAGGGAG 5972

Qy   961 GACCTGGGTTAACTTAACTTTACTCCCTGCCAATCCTCTTCGTGGTCGTGTCCCCTGCTGCGTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5973 GACCTGGGTTAACCTACTCCCTGCTATCCCTGCCAATCCCCTGGGTCGTGTCCCCTGCGTG 6032

Qy   1021 TGGCGCGCGATTACTCGGTCGGTGCACGGAGGGGAGGGAGGGTCGTGCAGTGGCAGGAAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6033 TGGCGCAGGGATACTGCGTTCGGCGCACCTGGCGCCAGGGAGGGGGCGTGTGCAGTGGAGATGAAC 6092

Qy   1081 CGGCTGATAGGGTTCGCCTCGCGGGTAACCAGTTTGCCCCCACTATGTGCCAGAG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6093 CGGGTCGATAGGGTTCGCCTTCGCGGGGGTAACCAGCGTCTCCCCACTATTGTGCCTTGAG 6152

Qy   1141 AGCGGACGCGCCGCAGCACGTGTCACTCAGATCCTCTTTCGCTGGTTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6093 CGGGCTGATAGGGTTCGCCTTCGCGGGGGTAACCAGCGTCTCCCCACTATTGTGCCTTGAG 6152

Qy   1141 AGCGGACGCGCCGCAGCACGTGTCACTCAGATCCTCTTTCGCTGGTTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6153 AGCGAGGACTGACGAGCACGTGTCACTCAGATCCTCTTTAGTCTTTACCAATCACTCAGCTGCTG 6212

Db 8373 CGCTTTACATCGGGGGCCCCTGACTAATTCTAAAAGGCAGAACTGCGGGGCTATCGCGCCG
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3421 TGCGGCGCCGAGCGGTGCGTGTTGCTGACGGACTAGCTGCGTAATACCTCACATGTTACTTGAAG 3480
Db 8433 TGCGCGCGGAGCGGTGACTAACCACACGGTAATACCTCACATGTTACTTGAAG 8492

XX
Qy 3481 GCCTCTGCAGCCGTGCGAGACTGCAAAAGTCTTCAGGAACTGACGATGCTCGTGTGCGGAGAC 3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8493 GCCCCTGCGCGCCCTGTCGACGAAGCTTCCAGGAACTGACGATGCTCGTTATGCGGGAGAC 8552

XX
Qy 3541 GGGCCTTGTTCGTTATTCTGTGGAGAGCGGGGGGAGGACCCAGGAACCCAGGACGCCTACGGAGTC 3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8553 GACCTTTGTCGTTATTCTGTGTGAAAGCGGGGGGACCCAAGAGACGACCCTACGGCGCC 8612

XX
Qy 3601 TTCACGGAGGCTATGACTAGGTACTCTGCCCCCCCCCCCCCCCCCCAACCAGAATAC 3660
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8613 TTCACGAGGGCTATGACTAGATACTCTGCCCCCCCCCCCCCTGGACCCCAAAACCAGAATAC 8672

XX
Qy 3661 GACCTGGAGTTGATAAACATCATGTCCTCCCTCCAAATGTTCGGTCGCGCGCCA CCGATGCATCTGGC 3720
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8673 GACTTGGAGTTGATAAACATCATGCTCCTCCCAAATGTGTCACTCGCCCACGATGCATCTGGC 8732

XX
Qy 3721 AAAAAGGGGTATACTACCTCACCCGTGACCC 3749
          ||||||||||||||||||||||||||||||
Db 8733 AAAAGGGGGTGTACTATCTCACCCGTGACCC 8761

RESULT 15
ABK91425

The invention relates to nucleic acid molecules comprising altered HCV
NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
internal ribosome entry site (IRES) region coding for one or more NS3,
NS5A, or EMCV IRES mutations, respectively. The location of the mutations
are detailed in the specification. Also included are (1) an expression
vector comprising a nucleotide sequence coding for the altered nucleic
acids, which is transcriptionally coupled to an exogenous promoter; (2) a
recombinant cell human hepatoma cell comprising the altered nucleic acids
; (3) a recombinant cell produced by introducing into a human hepatoma
cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
replicon enhanced cell or which containing a functional HCV replicon; (5)
an HCV replicon enhanced cells made by the method; and (6) measuring the
ability of a compound to affect HCV activity. The HCV replicons and HCV
replicon enhanced cells are useful in studying HCV replication and
expression, and HCV and host cell interactions, producing HCV RNA and
proteins, and providing a system for measuring the ability of a compound
to modulate one or more HCV activities e.g. to discover drugs which may
treat HCV mediated diseases such as liver failure, cirrhosis and
hepatocellular carcinoma. The present sequence is an HCV replicon Con 1
mutant of the invention. Note: The present sequence is not shown in the
specification but was created by the indexer using the HCV sequence
appearing as ABK91411 and the information in Claim 9

XX
SQ Sequence 9605 BP; 1909 A; 2883 C; 2734 G; 2079 T; 0 U; 0 Other;

Query Match   87.8%; Score 3291.4; DB 6; Length 9605;
Best Local Similarity   92.4%; Pred. No. 0;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

Qy   1 TGGGAGGGGGGTCTTCACAGCCTCACCCAGGTAGGGATGCCCACTTCCTGTCCCAAACAAG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5013 TGGGAGGAGGCGGTCTTTACAGGCCTCACCCAGGTAGGGATGCCCACTTCCTGTCGCTAGGCC 5072

Qy   61 CAGGCGAGGAGACACACTTCCCCTACCTGGGTGGCGTACCAGCGCTACTGTGTGCCGTAGGGC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5073 CAGGCGAGGAGCAAACTTCCCCTACCTGGTAGCATACCAGCGCTACGGTGTGCCCAGGGCT 5132

Qy   121 CAGGGCCCCACCTCCATCACTGGGGATCANAAATGTGGAAGGTGTCTCATACGGCGTAAAGCCTACT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5133 CAGGGCTCCACCTCCATCGGTGGGACCAAAATGTGGAAGGTGTCTCATACGGCGTAAAGCCTACG 5192

Qy   181 CTGCCGCGGGGCCCAACACCCCTTGCCTGTATAGGCTGGGAGCCCGTCCCAAAACGAGGGTCACCCTC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5193 CTGCCACGGGGCCCAACGCCCCTGCTGTPATAGGCCGGTCAAAACGAGGGTTACTACC 5252

Qy   241 ACACACCCCCATAACCAAAATTCATCANAGCCANTGGCATGGCTGGGAGGTCGCTCACG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   5253 ACACACCCCCATPAACCAAAATACATCANAGGCATGTCGGGCPAGGTCGTCACG 5312

Qy   301 AGCACCTGGGGTGCTCGGGTGCGGGGGGGTCCTTGCGAGCTCTGGCCTPATTGCTTGACAACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5313 AGCACCTGGGGTGCTCGGTGCGGAGTCTTCCTAGCCAGTCTTGGCCGCGGTPATTGCGTGACAACA 5372

Qy   361 GGCACGCGTGGGTCATTGTGGGTPAGGATCATCTTTGTCCGGGGGGGGCCCGGCCTPATTGTCCGGAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5373 GGCACGGCTGGGTCATTTGTGGGCPAGGATCATTCTTGTCCGGAAAGCCCGGCGCATCATTCCCGAC 5432

Qy   421 AGGGGAAGTCCTTCTACCAGGAGTTCGATGAGAGTGGGAGAGTTCGCGGTTCGCAACCTCCCTTAC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   5433 AGGGGAAGTCCTTTPACCGGGAGTTCGATGAGGAAGGGCGCGCCTTCACAACCTCCCTTAC 5492

Qy   481 ATCGAGCAGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAAGCGCCTCGGGTTGCTTGCAG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   5493 ATCGAACAGGGAAATGGCAGCTCGCCGAACAATTCAAACAGAAGGCAATCGGGCGTTGCTTGCAA 5552

Qy   541 ACAGCCACCAAGACGCGAGGAGGCCGCCTGCTCCCGTGTGGTGAGTCCAAGTGGGCAGCCCTT 600
           ||||||||||||||||||||||||||||||||||||||||||||
Db   5553 ACAGCCACCAAGAACGCGGAGCTCGCTGCGTCCCGTGGTGGPAATCCAAGTGGGCGGACCCTG 5612

Qy   601 GAGACCTTCTGGGGCGAAGCATPATGTGGAACTTCATCAGCGGATCAATATTAGCAGGC 660
           |||||||||||||||||||||||||||||||||||||||||
Db   5613 GAAGCCTTCTGGGCGAAGCATATGTCGGAATTTCATCAGCGGATACAATATTAGCAGGC 5672

```
Db   7833 AAGGCTAAACTTCTATCCGTGGGAGGAAGGCCTGTAAGCTGACGGCCCCCACATTCGGCCAGA 7892
          |||||||||||  |||||||||||||||||||  |||||||||||||||||||||||||||||| |
Qy   2881 TCTAAATTTGGGCTATGGGACGCAAAGGACGTCCGGAACCTATCCACCAAGGCCATTAACCAC 2940
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7893 TCTAAATTTGGCTATGGGGCGCAAAGGACGTCCGGAACCTATCCACGAAGGCCGTTAACCAC 7952

Qy   2941 ATCCGGCTCCGTGTGGGAGGACTTGTTGGAAGAGACACTGAAACACCAATTGACACACCATC 3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7953 ATCCGGCTCCGTGTGGGAAGGACTTCCTGGAAAGAGACACTGACACCAAACCATC 8012

Qy   3001 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGGAGGAGGGCCGCAAGCCAGCTCGC 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8013 ATGGCAAAAAATGAGGTTTTCTGGGTCCAACCAGAGGAGGGGCCGCAAGCCAGCTCGC 8072

Qy   3061 CTTATCGGTGTTCCCAGACTTGGGGGTCCTGTCCCGTGTGTGGCGAGAAAAATGGGCCCTCTATGACGTG 3120
          |||||||| |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8073 CTTATCGTTATTCCAGAATTTGGGGGTTCGTGTGTGTGCGAGAAAATGGCCCTTTACGGATGTG 8132

Qy   3121 GTCTCCACCCTCCCTCAGCGTGTGATGGGTCCTCGGTACGGGATTCCCAGGTATTCTCCTGA 3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8133 GTCTCCACCCTCCCTCAGCCCGTGATGGGGTCTTCATACGGGATTCCAAATACTCTCCTGA 8192

Qy   3181 CAGCGGGGTCGAGTTCCTGGTGAAACCCTGGAAATCAAAGGAGGACCCCTATGGGGGCTTTGCA 3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8193 CAGCGGGGTCGAGTTCCTGGTGAAATGCCTGGAAAGGGAGGAGAAAATGCCCTATGGGGGCTTCGGCA 8252

Qy   3241 TATGACACCCGCTGTTTTGACTCAACAGTCACTCGAGAATGACACTCCGTGTGAGAGAGTCA 3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8253 TATGACACCCGCTGTTTTTGACTCAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 8312

Qy   3301 ATTTATCAAATGTTCTGGACTTGGGCCCCGGAAGCCAGCAGGCCCATAAGGTCGGCTCACAGAG 3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8313 ATCTACCAAATGTTGTGACTTGGGCCCCCGGAAGGCCAGACAGGCCCATAAGGTCGCTCACAGAG 8372

Qy   3361 CGGGCTTTATATCGGGGGGGTCCCCTGACTAATTCAAAAGGGGCAGAACTGGCGGCTATGGCCGG 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8373 CGGGCTTTACATCGGGGGGCCCCCTGACTAATTCTAAAGGGCCAGAACTGCGGGCTATGGCCGG 8432

Qy   3421 TGCCGCGCGAGCGCGCGGTGCTGACGACTAGCTGCCGGTAAATACCCCTCACGTTACTTGAAG 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8433 TGCCGCGCGAGCGGGTGTACTGACCAGCTGCGGTAATACCCTCACATGTTACTTGAAG 8492

Qy   3481 GCCTCTGCGCAGCCTGCTGCGAGCTGCAAAAGCTTCCAGGACTGCACGGATGCTTCGTGTGCCGGAGAC 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8493 GCCGCGTGCGCCCTGTCGCCGAAGCCTCCAGAGACTGCACGGATGCTTCGTATGCCGGAGAC 8552

Qy   3541 GGGCCTTGTCGTTATCTGTGAGAGCGCGGAACCCAGGAGGAGGGGGGAGCCTACCGAGTC 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8553 GACCTTGTCGTTATCTGTGAAAGCGCGGGAAAGGAGGAGGGGACGAGCCCTACCGGGGCC 8612

Qy   3601 TTCACGGAGCGCTATGACTAGGTATACTCTGCCCGCCCCCCGGGGGGACCCCCCAAACCAGAATAC 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8613 TTCACGGAGGCTATGACTAGATACTTCTGCCCCCCCCTGGGGGACCCGGCCAAAACCAGGATAC 8672

Qy   3661 GACCTGGGAGTTGAATAAACATCAATGCTCCTCCAAATGTTGTCGGTCGCGGCACGACGATGCATCTGGC 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8673 GACTTGGGAGTTGATAAACATCATGCTCCTCCAAATGTTCAGTGCCGCCACGATGCACTCTGGC 8732

Qy   3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
          |||||||| |||||||| |||||||||||||
Db   8733 AAAAGGGTGTACTATCTCACCCGTGACCC 8761
```

```
OM nucleic - nucleic search, using sw model
                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:             February 21, 2005, 10:33:33   ; Search time 584 Seconds
                                                    (without alignments)
                                                    10506.914 Million cell updates/sec

Title:              US-09-664-363-22
Perfect score:      3750
Sequence:           1 TGGGAGGGGGCTCTTCACAGG............ACTACTCACCGGTGACCCG 3750
Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0
Searched:           1202784 seqs, 81813B359 residues

Total number of hits satisfying chosen parameters:     2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :          Issued Patents NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3750	100.0	3750	3	US-08-191-160-22	Sequence 22, Appl
2	3292.2	87.8	8642	4	US-10-029-907-2	Sequence 2, Appl
3	3291.4	87.8	7989	4	US-09-539-601-10	Sequence 10, Appl
4	3291.4	87.8	8001	4	US-09-539-601-7	Sequence 7, Appl
5	3291.4	87.8	8637	4	US-09-539-601-4	Sequence 4, Appl
6	3291.4	87.8	8638	4	US-10-029-907-24	Sequence 24, Appl
7	3291.4	87.8	8639	4	US-10-029-907-1	Sequence 1, Appl
8	3291.4	87.8	8649	4	US-09-539-601-13	Sequence 13, Appl
9	3291.4	87.8	11076	4	US-09-539-601-1	Sequence 1, Appl
10	3288.2	87.7	8638	4	US-10-029-907-6	Sequence 6, Appl
11	3288.2	87.7	8638	4	US-10-029-907-7	Sequence 7, Appl
12	3288.2	87.7	8638	4	US-10-029-907-25	Sequence 25, Appl
13	3288.2	87.7	8643	4	US-10-029-907-4	Sequence 4, Appl
14	3285	87.6	8648	4	US-10-029-907-5	Sequence 5, Appl
15	3281.8	87.5	8001	4	US-09-539-601-22	Sequence 22, Appl
16	3281.8	87.5	11076	4	US-09-539-601-25	Sequence 25, Appl
17	3280.2	87.5	8001	4	US-09-539-601-16	Sequence 16, Appl
18	3280.2	87.5	11076	4	US-09-539-601-19	Sequence 19, Appl
19	3278.6	87.4	7863	1	US-08-324-977-35	Sequence 35, Appl
20	3278.6	87.4	7863	2	US-08-384-616-35	Sequence 35, Appl
21	3278.6	87.4	7863	2	US-08-904-686A-35	Sequence 35, Appl
22	3278.6	87.4	7863	4	US-09-315-850-35	Sequence 35, Appl
23	3278.6	87.4	7917	1	US-08-324-977-31	Sequence 31, Appl
24	3278.6	87.4	7917	2	US-08-384-616-31	Sequence 31, Appl
25	3278.6	87.4	7917	2	US-08-904-686A-31	Sequence 31, Appl
26	3278.6	87.4	7917	3	US-09-315-850-31	Sequence 31, Appl
27	3278.6	87.4	9030	1	US-08-324-977-13	Sequence 13, Appl
28	3278.6	87.4	9030	2	US-08-384-616-13	Sequence 13, Appl
29	3278.6	87.4	9030	2	US-08-904-686A-13	Sequence 13, Appl
30	3278.6	87.4	9030	3	US-09-315-850-13	Sequence 13, Appl
31	3278.6	87.4	9416	1	US-08-324-977-1	Sequence 1, Appl
32	3278.6	87.4	9416	2	US-08-384-616-1	Sequence 1, Appl
33	3278.6	87.4	9416	2	US-08-904-686A-1	Sequence 1, Appl
34	3278.6	87.4	9416	3	US-09-315-850-1	Sequence 1, Appl
35	3278.6	87.4	9416	4	US-09-827-688A-27	Sequence 27, Appl
36	3275.4	87.3	8001	4	US-08-823-895A-27	Sequence 28, Appl
37	3275.4	87.3	11076	4	US-09-539-601-31	Sequence 31, Appl
38	3230.6	86.1	9595	4	US-09-014-416-4	Sequence 4, Appl
39	3200.6	85.3	9413	4	US-09-827-688-6	Sequence 6, Appl
40	3192.2	85.1	9472	4	US-08-150-204E-96	Sequence 96, Appl
41	2449.6	65.3	12980	4	US-08-811-566-5	Sequence 5, Appl
42	2449.6	65.3	12980	3	US-09-034-756-5	Sequence 5, Appl
43	2448	65.3	9646	3	US-08-811-566-1	Sequence 1, Appl
44	2448	65.3	9646	3	US-09-034-756-1	Sequence 1, Appl
45	2446.4	65.2	9365	4	US-09-827-688-7	Sequence 7, Appl

```
                                ALIGNMENTS

RESULT 1
US-08-191-160-22
Sequence 22, Application US/08191160
Patent No. 6210675
GENERAL INFORMATION:
 APPLICANT: Highfield, Peter Edmund
 APPLICANT: Rodgers, Brian Colin
 APPLICANT: Tedder, Richard Seton
 APPLICANT: Barbara, John Anthony James
 TITLE OF INVENTION: Viral Agent
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
  ADDRESSEE: Rothwell, Figg, Ernst & Kurz
  STREET: 1700 K Street
  CITY: Washington
  STATE: D.C.
  COUNTRY: U.S.A.
  ZIP: 20006
 COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy diskette, 5.25 inch, 360 Kb storage
  COMPUTER: IBM AT compatible
  OPERATING SYSTEM: MS-DOS V3.2
  SOFTWARE: Wordperfect 5.0 (DOS text)
 CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/191,160
  FILING DATE:
 PRIOR APPLICATION DATA:
  CLASSIFICATION:
 PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 07/628,516
  FILING DATE: 17 DEC 1990
 PRIOR APPLICATION DATA:
  APPLICATION NUMBER: UK 89 28 562.1
  FILING DATE: 18 DEC 1989
 PRIOR APPLICATION DATA:
  APPLICATION NUMBER: UK 90 414.0
  FILING DATE: 27 FEB 1990
 PRIOR APPLICATION DATA:
  APPLICATION NUMBER: UK 90 04 814.1
  FILING DATE: 03 MAR 1990
 ATTORNEY/AGENT INFORMATION:
  NAME: E. Anthony Figg
  REGISTRATION NUMBER: 27,195
  REFERENCE/DOCKET NUMBER: 1645-103A
 TELECOMMUNICATION INFORMATION:
  TELEPHONE: (202) 833-5740
  TELEFAX: (202) 833-5744
 INFORMATION FOR SEQ ID NO: 22:
  SEQUENCE CHARACTERISTICS:
   LENGTH: 3750 base pairs
   TYPE: nucleotide with corresponding protein
   STRANDEDNESS: single
```

```
;    TOPOLOGY:    linear
;    MOLECULE TYPE:    cDNA to genomic RNA
;    ORIGINAL SOURCE:
;      ORGANISM:    human; serum infectious for PT-NANBH
;      IMMEDIATE SOURCE:
;      LIBRARY:    cDNA clones from 3' end of the genome
;    FEATURE:
;      LOCATION:    from 1 to 3750 bp portion of the PT-NANBH
;      LOCATION:    polyprotein
;      OTHER INFORMATION:    viral non-structural proteins

US-08-191-160-22

Query Match    100.0%;    Score 3750;    DB 3;    Length 3750;
Best Local Similarity    100.0%;    Pred. No. 0;    Mismatches 0;
Matches 3750;    Conservative    0;    Mismatches 0;    Indels 0;    Gaps 0;
```

Qy   1 TGGGAGGGGGGTCTTCACAGCCCTCACCCCAGTGGATGCCCACTCCTTGTCCCAAAACAAAG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TGGGAGGGGGGTCTTCACAGCCCTCACCCCAGTGGATGCCCACTCCTTGTCCCAAAACAAAG 60

Qy   61 CAGGCAGGAGAGACAACTTCCCCTACCTGGTGGGCGTACCAGGCGTACTGTGTTAGGGGCC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAGGCAGGAGAGACAACTTCCCCTACCTGGTGGGCGTACCAGGCGTACTGTGTTAGGGGCC 120

[The page continues with a full DNA sequence alignment of Query (Qy) and Database (Db) sequences running from position 1 to approximately 1980, with contiguous identity match bars between each pair of rows. The alignment consists of paired 60-base-pair lines labeled Qy and Db with vertical match indicators throughout, representing a 100% identity match over the full 3750 bp length.]

```
Db  1921  TTGCCAGCTCTTCTTCAGCTTAGTTGTCTGCGCGCCTTTCCTCGGAAGGCGACATACATTACC  1980

Qy  1981  CAAAATGACTTCCCAGACCGCTGACCTCATCGAGGCCAACCTCCTGTTGTGGCGGCCATGAGATG  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  CAAAATGACTTCCCAGACCGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGGCCATGAGATG  2040

Qy  2041  GGCGGGGACATTACCCGCCGTGGAGTTCAGAGAACAAAGGTAGTAATCCTCTTCTTTCGAC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  GGCGGGGACATTACCCGCCGTGGAGTTCAGAGAACAAAGGTAGTAATCCTCGTGGCCTCTTTCGAC  2100

Qy  2101  CCCGCTCCGAGCGCGAGGCGGATGACGGGAAGTGTCGGTCCCGGCGGGAGATCCTGCGGGAAA  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  CCCCTCCGAGCGCGGAGGATGAGGGGAAGTGTCCGTCCCGGCGGGAGATCCTGCGGGAAA  2160

Qy  2161  TCCAAGAAATTCCCACCACAGGGCATGGGCCGCGCCGGATTACAACCCTCCGGCTG  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  TCCAAGAAATTCCCACCACAGGGCATGGGCCGCGCCGGATTACAACCCTCCGGCTG  2220

Qy  2221  CTGGAGTCCTGGAAGGCCCCGGACTACGGTCCCTCCCAGTGGTGTACATGGGGTCCCACTGCCA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  CTGGAGTCCTGGAAGGCCCCGGACTACGGTCCCTCCCAGTGGTGTACATGGGGTCCCACTGCCA  2280

Qy  2281  CCTACTAAGACCCCTCCTATACCACCTCACACTCCACGGAGGAGGACAGTTGTTCTGACAGAA  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  CCTACTAAGACCCCTCCTATACCACCTCACACTCCACGGAGGAGGACAGTTGTTCTGACAGAA  2340

Qy  2341  TCCACCCGTGTCTTCTGCCCTGGGCGGAGCTTGCCCGAGCTTGCCGACGGGGTCCGAAACCG  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TCCACCCGTGTCTTCTGCCCTGGGCGGAGCTTGCCCGAGCTTGCCGACGGGGTCCGAAACCG  2400

Qy  2401  TCGGCCGGTCGACAGCGGCACGGCGACCGCCCGTCCTCTGACCACCCCTCGACGACGGGCGGA  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TCGGCCGTCGACAGCGGCACGGCGACCGCCCGTCCTCTGACCACCCCTCGACGACGGGCGGA  2460

Qy  2461  GCAGGATCTGACGTTGAGTCGGTAATTCCTTCAATGCCCGTGCCCCCTTTGAGGGGGGGGAGCCCGGGGAC  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GCAGGATCTGACGTTGAGTCGGTATTCCTCCATGCCCGTGCCCCCTTTGAGGGGGGGGAGCCCGGGGAC  2520

Qy  2521  CCCGATCTCAGCGACGGGGTCTTGGTCTACGGTGAGTGAGGGCGGTGAGGACGTCGTC  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CCCGATCTCAGCGACGGGGTCTTGGTCTACGGTGAGTGAGGGCGGTGAGGACGTCGTC  2580

Qy  2581  TGCTGCTCGGATGTCCTCACACATGACACGGCGCTCTGATCACGGCCATGTCGCGGAGGAA  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TGCTGCTCGGATGTCCTCACACATGACACGGCGCTCTGATCACGGCCATGTCGCGGAGGAA  2640

Qy  2641  AGCAAGCTGCCCATCAAACGGCGTTGAGCAACTCTTTGCTGCGGTCACCACAACACATGGTCTAC  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  AGCAAGCTGCCCATCAAACGGCGTTGAGCAACTCTTTGCTGCGGTCACCACAACACATGGTCTAC  2700

Qy  2701  GCTACCACATCCCGCAGCGCCAAGCCCAGCGCCAGAGAAGGTCACCTTTGACAGACTGCAA  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  GCTACCACATCCCGCAGCGGCAAGCCCAGCGCCAGAGAAGGTCACCTTTGACAGACTGCAA  2760

Qy  2761  ATCCTGGACGACTACACTACCAGGACGACGTGCTCAAGCGAGGATGAAGGCCAAGGCGTCCACAGTT  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  ATCCTGGACGACTACACTACCAGGACGACGTGCTCAAGCGAGGATGAAGGCCAAGGCGTCCACAGTT  2820

Qy  2821  AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCAGCGCCCCACATTCGGCCAAA  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCAGCGCCCCACATTCGGCCAAA  2880

Qy  2881  TCTAAATTTGGCTATGGGCGGGACACGTCCGGGACCTATCCAGCGAAGGCCATTAACCAC  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  TCTAAATTTGGCTATGGGCGGGACACTGTTGGGAAAGACGTCCGGGACCTATCCAGCGAAGGCCATTAACCAC  2940

Qy  2941  ATCCGGCTCCGTGTGTGGGGAGGACTTGTTCGGAAAGACACTGAAAACACCAATTGACACCACCATC  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  ATCCGGCTCCGTGTGTGGGGAGGACTTGTTCGGAAAGACACTGAAAACACCAATTGACACCACCATC  3000

Qy  3001  ATGGCAAAAAATGAGGGTTTTCTGGGTCCAACCAGAGAGGAGGAGGGCCGCGAAGCCAGCTCGC  3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  ATGGCAAAAAATGAGGGTTTTCTGGGTCCAACCAGAGAGGAGGAGGGCCGCGAAGCCAGCTCGC  3060

Qy  3061  CTTATCGTGTTCCCAGACTTGGGGCGGTCCGGTGTCGTGTGTGGAGAAAATGGCCCTCTATGACGTG  3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  CTTATCGTGTTCCCAGACTTGGGGCGGTCCGGTGTCGTGTGTGGAGAAAATGGCCCTCTATGACGTG  3120

Qy  3121  GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTGTTGGTTACGGATTCCAGTATTCTCCTGGA  3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTGATGGGGCTCCTCGTGTACCGGATTCCAGTATTCTCCTGGA  3180

Qy  3181  CAGCGGGGTCGAGTTCCTGTGGTGAAACGGCCTGGAAATCAAGAAGAGACCCCTATGGGGCTTTGCA  3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  CAGCGGGGTCGAGTTCCTGTGGTGAAACGGCCTGGAAATCAAGAAGAGACCCCTATGGGGCTTTGCA  3240

Qy  3241  TATGACACCCCCTGTTTTTGACTCAACAGTCACTGAGAAATGACATCGTGTAGAGGAGTCA  3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TATGACACCCCCTGTTTTTGACTCAACAGTCACTGAGAAATGACATCGTGTAGAGGAGTCA  3300

Qy  3301  ATTTATCAATGTTGTGACTTGGCCCCCGAAGCCAGCCAGACAGCAGGCCATAAGGTCGGCTCACAGAG  3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  ATTTATCAATGTTGTGACTTGGCCCCCGAAGCCAGCCAGACAGCAGGCCATAAGGTCGGCTCACAGAG  3360

Qy  3361  CGGGCTTTATTATCGGGGGGTCCCCTGACTAATTCAAAGGGGGGAACTGCGGCGTACGCCGG  3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  CGGGCTTTATTATCGGGGGGTCCCCTGACTAATTCAAAGGGGGGAACTGCGGCGTACGCCGG  3420

Qy  3421  TGCCGCGGCGAGCGGGGTGCGTGACGACTAGCTGCGGTAATACCCTCACACATGTTACTTGAAG  3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  TGCCCGCGGCGAGCGGGGGTGCGTGACGACTAGCTGCGGTAATACCCTCACACATGTTACTTGAAG  3480

Qy  3481  GCCTCTGCAGCCTGTCGGAGCTTGCAAAGCTCCAGGACTGCACGACAGATGCTCGTGTGTGCGGAGAC  3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  GCCTCTGCAGCCTGTCGGAGCTTGCAAAGCTCCAGGACTGCACGACAGATGCTCGTGTGTGCGGAGAC  3540

Qy  3541  GGGCCTTGTCGTTATCTGTGTGAGAGCGCGGGAAACCCAGGAGGAGGAGGGCGGCCTAGGAGTC  3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  GGGCCTTGTCGTTATCTGTGTGAGAGCGCGGGAAACCCAGGAGGAGGAGGGCGGCCTAGGAGTC  3600

Qy  3601  TTCACGGAGGGCTATGACTAGGTACTCTGCCCCCCCGCCCCCCAACCAGAATAC  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  TTCACGGAGGGCTATGACTAGGTACTCTGCCCCCCCGCCCCCCAACCAGAATAC  3660

Qy  3661  GACCTGGAGTTGATAACATCATCTCCTCCAAATGTGTCGTGTCGGCGCCGACGATGCATCTGGC  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  GACCTGGAGTTGATAACATCATCTCCTCCAAATGTGTCGTGTCGGCGCCGACGATGCATCTGGC  3720

Qy  3721  AAAAGGGTATACTACCTCACCCGTGACCCG  3750
          ||||||||||||||||||||||||||||||
Db  3721  AAAAGGGTATACTACCTCACCCGTGACCCG  3750
```

```

```
;   LOCATION: 6268
;   OTHER INFORMATION: r = a or g
;   NAME/KEY: variation
;   LOCATION: 4446
;   OTHER INFORMATION: r = a or g
US-10-029-907-2

  Query Match   87.8%; Score 3292.2; DB 4; Length 8642;
  Best Local Similarity   92.3%; Pred. No. 0;
  Matches 3462; Conservative   2; Mismatches 285; Indels   0; Gaps   0;

Qy   1 TGGGAGGGCGTCTTCACAGGCCTCACCCACGTGGATGCCCACTTCCTGTCCCAAACAAAG 60
                |||||| ||||||| ||||||||||||||| | || ||||| ||| ||||||| || |||
Db   4046 TGGGAGAGCGTCTTTACAGGCCTCACCCACATAGACGCCCATTTCTTGTCCCAGACTAAG 4105

Qy   61 CAGGCAGGAGACAACTTCCCCTACCTGGTGGCGTACCAGGCTACTGTGTGCGCTAGGGCC 120
                ||||||||||||||||||||||||||||| || ||||||||||| |||||||| |||||
Db   4106 CAGGCAGGAGACAACTTCCCCTACCTGGTAGCATACCAGGCTACGGTGTGCGCCAGGGCT 4165

Qy   121 CAGGCCCCACCTCCATCATGGGATCAAATGTGGAAGTGTCTCATACGGCTAAAGCCTACT 180
                ||||| ||||||||||| ||||| |||||||||||||||||||||||||||||||||||
Db   4166 CAGGCTCCACCTCCATCGTGGGACCAAATGTGGAAGTGTCTCATACGGCTAAAGCCTACG 4225

Qy   181 CTGCGCGGGCCAACACCCTTGCTGTATAGGCTGGGAGCCGTCCAAAACGAGGTCACCCTC 240
                |||| ||||||||| ||| |||||||||||||||||||||| ||||||||||| || |
Db   4226 CTGCACGGGCCAACGCCCCTGCTGTATAGGCTGGGAGCCGTTCAAAACGAGGTTACTACC 4285

Qy   241 ACACACCCCATAACCAAATTCATCATGGCATGCATGTCAGCCGACCTGGAGGTCGTCACG 300
                ||||||||||||||||||| |||||||||||||||||| || ||||||||||||||||||
Db   4286 ACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTCACG 4345

Qy   301 AGCACCTGGGTGCTGGTGGGCGGGGTCCTTGCAGCTCTGGCTGCGTATTGCTTGACAACA 360
                ||||||||||||||||| ||||| ||||| ||||||||||| ||||||||| ||||||||
Db   4346 AGCACCTGGGTGCTGGTAGGCGGAGTCCTAGCAGCTCTGGCCGCGTATTGCCTGACAACA 4405

Qy   361 GGCAGCGTGGTCATTGTGGGTAGGATCATCTTGTCCGGGCGGCCGGCTATTGTTCCCGAC 420
                |||||||||||||||||||| ||||||||||||||||| :|||||| || ||||||||
Db   4406 GGCAGCGTGGTCATTGTGGGCAGGATCATCTTGTCCGGAARGCCGGCCATCATTCCCGAC 4465

Qy   421 AGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCGCGTCGCACCTCCCTTAC 480
                ||||||||||| |||| ||||||||||||||||||||||||||| || ||||||||||||
Db   4466 AGGGAAGTCCTTTACCGGGAGTTCGATGAGATGGAAGAGTGCGCCTCACACCTCCCTTAC 4525

Qy   481 ATCGAGCAGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAGCGCTCGGGTTGCTGCAG 540
                ||||| |||||||||||||||||||| || ||||| || || || |||||||||||||
Db   4526 ATCGAACAGGGAATGCAGCTCGCCGAACAATTCAAACAGAAGGCAATCGGGTTGCTGCAA 4585

Qy   541 ACAGCCACCAAGCAAGCGGAGGCCGCTGCTCCCGTGGTGGAGTCCAAGTGGCGAGCCCTT 600
                ||||||||||||||||||||||| ||||||||||||||||| ||||||||||| ||||
Db   4586 ACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCGTGGTGGAATCCAAGTGGCGGACCCTC 4645

Qy   601 GAGACCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTACTTAGCAGGC 660
                || ||||||||||||| || |||||||| ||||||||||||||||| || |||||||||
Db   4646 GAAGCCTTCTGGGCGAAGCATATGTGGAATTTCATCAGCGGGATACAATATTTAGCAGGC 4705

Qy   661 TTGTCCACTCTGCCTGGGAATCCCGCGATTGCATCACTGATGGCGTTCACAGCCTCTGTC 720
                ||||||||||||||||| || |||||||| |||||||||||||| |||||||||||| ||
Db   4706 TTGTCCACTCTGCCTGGCAACCCCGCGATAGCATCACTGATGGCATTCACAGCCTCTATC 4765

Qy   721 ACTAGCCCGCTCACCACCCAATCTACCCTCCTGCTTAACATCCTGGGGGGATGGGTAGCC 780
                || |||||||||||||||||| |||||||||| |||||||||||||||||||||| |||
Db   4766 ACCAGCCCGCTCACCACCCAACATACCCTCCTGTTTAACATCCTGGGGGGATGGGTGGCC 4825

Qy   781 GCCCAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTAGGCGCCGGCATTGCTGGTGCG 840
                |||||||| ||||| ||||| |||||||| |||||||||||||||||||| ||||| |||
Db   4826 GCCCAACTTGCTCCTCCCAGCGCTGCTTCTGCTTTCGTAGGCGCCGGCATCGCTGGAGCG 4885

Qy   841 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGACATCTTGGCGGGCTATGGAGCA 900
                |||||||||||||||||||||||||||||||||||||| || ||||| || |||||||||
Db   4886 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGATATTTTGGCAGGTTATGGAGCA 4945

Qy   901 GGAGTGGCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGGCGAAATGCCCTCCACCGAG 960
                || ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   4946 GGGGTGGCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACCGAG 5005

Qy   961 GACCTGGTTAACTTACTCCCTGCCATCCTCTCTCCTGGTGCCCTGGTCGTCGGGGTCGTG 1020
                |||||||||||| |||||||||| |||||||| ||||| ||||| |||||||||||||||
Db   5006 GACCTGGTTAACCTACTCCCTGCTATCCTCTCCCCTGGCGCCCTAGTCGTCGGGGTCGTG 5065

Qy   1021 TGCGCAGCGATACTGCGTCGGCACGTGGGTCCAGGGGAGGGGGCTGTGCAGTGGATGAAC 1080
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   5066 TGCGCAGCGATACTGCGTCGGCACGTGGGCCCAGGGGAGGGGGCTGTGCAGTGGATGAAC 5125

Qy   1081 CGGCTGATAGCGTTCGCCTCGCGGGGTAACCATGTTTCCCCCACGCACTATGTGCCAGAG 1140
                ||||||||||||||||| |||||||||||||| |||||||||||||||||||||| |||
Db   5126 CGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCTGAG 5185

Qy   1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTCTCCGACCTTACTATCACCCAACTGTTG 1200
                |||||||| |||||||||||||||||||||||||| ||||| ||||| || ||| ||
Db   5186 AGCGACGCTGCAGCACGTGTCACTCAGATCCTCTCTAGTCTTACCATCACTCAGCTGCTG 5245

Qy   1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCACGCCCTGCTCCGGCTCGTGGCTA 1260
                |||||||| ||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db   5246 AAGAGGCTTCACCAGTGGATCAACGAGGACTGCTCCACGCCATGCTCCGGCTCGTGGCTA 5305

Qy   1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
                || ||||||||||| ||||||||||| || ||| |||| |||||||||||||||||||||
Db   5306 AGAGATGTTTGGGATTGGATATGCACGGTGTTGACTGATTTCAAGACCTGGCTCCAGTCC 5365

Qy   1321 AAGCTCCTGCCGCGATTACCGGGAGTCCCCTTTTTCTCATGCCAACGTGGGTACAAGGGG 1380
                ||||||||||||||||| |||||||||||||| |||||||| |||||||||||||||||
Db   5366 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 5425

Qy   1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440
                ||||||||||| |||||||||||||| ||||||||| |||||||||||||||||||||||
Db   5426 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 5485

Qy   1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
                ||||| |||||| |||||||||||||||| ||||||| |||||||||||||| |||||||
Db   5486 CATGTGAAAAACTGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5545

Qy   1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560
                |||||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||
Db   5546 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 5605

Qy   1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5606 TATTCTAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 5665

Qy   1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACGTAAAATGCCCGTGCCAGGTTCCA 1680
                |||||||||||||||||| |||||||||||||||||||||| |||||||| ||||||||
Db   5666 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5725

Qy   1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTGC 1740
                |||||||||||||||||||||||||||||||||||| |||||||||||||||| ||||||
Db   5726 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGCTCCAGCGTGC 5785

Qy   1741 AAACCTCTCCTACGGGAGGAGGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTTGGG 1800
                ||||| ||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db   5786 AAACCCCTCCTACGGGAGGAGGTCACATTCCTGGTCGGGCTCAATCAATACCTGGTTGGG 5845

Qy   1801 TCGCAGCTCCCATGCGAGCCCGAACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 1860
                || |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   5846 TCACAGCTCCCATGCGAGCCCGAACCGGACGTAGCAGTGCTCACTTCCATGCTCACCGAC 5905

Qy   1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920
                ||||||||||| || || |||||||||||||| |||||||||||||| ||||||||||||
Db   5906 CCCTCCCACATTACGGCGGAGACGGCTAAGCGTAGGCTGGCCAGGGGATCTCCCCCCTCC 5965

Qy   1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980
                ||||||||||| |||||||||||| ||||||||||||||| |||||| |||| || ||||
Db   5966 TTGGCCAGCTCATCAGCTAGCCAGCTGTCTGCGCCTTCCTTGAAGGCAACATGCACTACC 6025

Qy   1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040
                | |||||| ||| |||||||||||||||||||||||||||||||||||||| ||||||
Db   6026 CGTCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATG 6085
```

Qy 2041 GGGGGGGACAGCATTACCCGGGTGTGGGAGTCAGGAGAACAAGGTTAGTAATCCTTGGACTCTTTTCGAC 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6086 GGGGGGGACAGCATTACCCGGGTGTGGGAGTCAGAAAATAAGGTAGTAATTTTGGACTCTTTTCGAG 6145

Qy 2101 CCGGCTCCGAGCGGGAGGAGGGAGGATGAGGGGGAAGTGTCCGTCCGGGGGGGAGATCCTGCGGGAAA 2160
          |||||||||||||||||||||||||||||||||||||||
Db 6146 CCGGCTCCAAGCGGGAGGAGGGATGAGGGGGAAGTATCCGTTCCGGGGGGGAGATCCTGCGGGAGG 6205

Qy 2161 TCCAAGAAAATTCCCACCAGCGGATGCCCCGCGGCGGCCCCGGGATTACAAACCCTCCGGCTG 2220
          |||||||||||||||||
Db 6206 TCCAGGAAAATTCCCTGACGGATGCCCCATATGGGCACGGCCGCGCGGCGGCCCCGGG 6265

Qy 2221 CTGGAGTCCTGGAAGGCCCCCGGGGACTACGGTCCCCTTCCAGTGGCTGTACATGGGTGCCCACTGCCA 2280
          |:|||||||||||||||||||||||||||||||||||||
Db 6266 TTRGAGTCCTTGGAAGGACCCGGGACTACGTCCCCTGCCAGTCATTGCGG 6325

Qy 2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAGAGAGGACAGTTGTTCTGACAGAA 2340
          |||||||||||
Db 6326 CCTGCCAAAGGCCCCTCCGGATACCACCTCCGGGAGGAGAGAGGGCATCGGCCTATGCCGG 6385

Qy 2341 TCCACCCGTGTCTTTCTTGCCCTGGCACAAAGGCTTTTCGGTTAGCTTCGGAACCG 2400
          |:||||||||||||||||||||
Db 6386 TCTACCCGTGTCTTCTGCCTTGGCGGAGCTGCCACACTTCCAGTGGCTGCCC 6445

Qy 2401 TCGGCCCGTTCGCAGACCGGCACACGGCAACCGGCCCTTCCTGACCCTCCGGACGAGGAGGAC 2460
          ||||||||||||||||||||||||
Db 6446 TCCGCCCGTTCGAGCGGCACGGGGTCTTGGTCTACCGTAAGCGAGGAGGCTGGTGGGAACCTGG 6505

Qy 2461 GCAGGAGATCTGACGTTGAGTCGTGTATTCCTCCAGGTGTGGGGTAATCCTCACAGTGGTTAGGAAG 2520
          |||||
Db 6506 GCGGGGATCCGATCCGGGGAGTGGCCGGGCGGCGGGGAGGCCGGGGAT 6565

Qy 2521 CCCGATCTCAGCGGAGCAGCGGGTCTTGGTCTACCGTGAGTGTGAGGAGGTCGTC 2580
          ||||||||||||||
Db 6566 CCCGGATCTCAGCGACGGGTCTGGTCTACCGTAAGCGAGGAGGCTGGTGGGAA 6625

Qy 2581 TGCTGCTTCGGAGTGTCCTACACATGGACCAGGCGGTCTGATCAGCCATGCCCTGCCGGAGGAA 2640
          ||||
Db 6626 TGCTGCTTCGGATGTCCTTCACAGGACGAGCATGCCCTGATGCCATGTGCTGCGGGAGGAA 6685

Qy 2641 AGCAAGCTGCCCATCAACCCGGTTGAGCAACTCTTTTGCTCGTTCTCGTCACCAACATGGTCTCTAC 2700
          ||:||||||||||||
Db 6686 ACCAAGCTGCCCCATCAAATGCACTGAGCAACTCTTTTGCTCCGTCACCAACAACTTGGTCTTAT 6745

Qy 2701 GCTACCCACCATCCCGCAGGCGCAAGCCCAGCGGCCGCGAAGAGGGTCACCTTTTGACGACTGCGAA 2760
          |||||||||||||||||||||||||||||||
Db 6746 GCTACAACATCTCGCAGCCCAAGCCTGCGGCAGCGAAGAAGGTCACCTTTTGCAGAGACTGCAG 6805

Qy 2761 ATCCTGGACGAGATCACTACCAGGAGACGTGCTTCAAGGAGATGAAAGGGCAGGGCTTCCACAGTT 2820
          ||:||||||||||||||
Db 6806 GTCCTGGACGACCACTACGGGACGGGTCTTCAAGGAGAGTGAAAGGGGGCAAGGCGGTTCCACAGTT 6865

Qy 2821 AAGGCTAAGCTTCTTATCAGTAGAGAGAGCGCTGCACGGCGAGTCGACGCGCCGCCCACACTTCGGCCAAA 2880
          |||||||||
Db 6866 AAGGCTAAACTTCTTATCCGTGGAGGAAGCCTGTAGCGCTGACGCCCCCACATTCGGCCATCGGC 6925

Qy 2881 TCTAAAATTTGGGCTATTGGGGGACAGCGTCCCGGAACCTTATCCAGCAGGCCCATTAAACCAC 2940
          |||||
Db 6926 TCTAAATTTGGCTATTGGGGCACGTTCCGGAACCTATCCAGCAAGGCCGTTAACCAC 6985

Qy 2941 ATCCGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000
          ||||||||
Db 6986 ATCCGCTCCGTGTGGGAAGGACTTGCTCTGGAAGACAGCACCACCAGTGTAG 7045

Qy 3001 ATGGCAAAAAATGAGGGTTTTTCTGGCGTCCAACCACAGCAGAGAGGCCGCAAGCCAGCCTCGC 3060
          |||||||||||||
Db 7046 ATGGCAAAAAATGAGGTTTTTCTGGCGTCCAACCACAGAGAGAAGGGCCGCAAGCCCAGCTCGC 7105

Qy 3061 CTTATCGTGTTCCCAGACTTGGGGGGTCCGTGTGTGCGAGGAAAAATGGCCCTTTACCGATGTG 3120
          ||||||||||
Db 7106 CTTATCGTGTTCCCAGATTTTGGGGGGTTCGTGTGTCGGAGAAAATGGCCCTTTACGGATGTG 7165

Qy 3121 GTCTCCACCCTCCCCTCAGGGCTGTGGGATGGGTCCTCGTGTACGGATTCCAGTGTATTCTCCTGGA 3180
          |||||||||||||||||||
Db 7166 GTCTCCACCCTCCCCTCAGGGCCGTGGGGGCTCTTCATCAGGATTCCAATACTCCTGGA 7225

Qy 3181 CAGCGCGGTCGAGTTCCTGGGTGAACGCCCTGGAAATCAAAGAAGAACCCCTATGGGGCTTGCA 3240
          ||||||||||||||||||||||||||||
Db 7226 CAGCGCGGTCGAGTTCCTCGGTGAATCCCTGGAAAAGCGGAAGAAAATGCCCTATGGGGCTTCGGC 7285

Qy 3241 TATGACACCCGGCGTTTTGGACTCAACAGTCACTGAGAATGACATCCGGTTGTTGAGAGAGTCA 3300
          |||||||||||||||||
Db 7286 TATGACACCCGGCGTTTTTCAACGGGTCAGCTGAGAATGACATCCGGTTGAGAGAGTCA 7345

Qy 3301 ATTTATCAAATGTTGTGACTTGGCCCCGAGCCCGAGCCAGACCAGGCCGCCATAAGGTGCTCACAGAG 3360
          ||||||||||||||
Db 7346 ATCTACCAAATGTTGTGACTTGGCCCCGGAGCCCGCAGCACCGCCATAAGGGTCACTGGAG 7405

Qy 3361 CGGGCTTTATTATCGGGGGGGTCCCTGACTAATTCAACAGTGACACTGAGGAACTGCCGGCTATGCCGG 3420
          |||||||||||
Db 7406 CGGGCTTTACATCGGGGGGGGCCCCTGACTAATTCTAAAAGGGCAGAAACTGCCGGCTATGCCCGG 7465

Qy 3421 TGCCCGGGCGAGCGGGTGCTGACGGACTAGGTGCGGTAATAACCCTCACAGTGTTGTGAAG 3480
          |||||||||||||||
Db 7466 TGCCGCGGCGAGCGGGTGAGTACTGGACGAGGCCAGGTAATACCCTCACAGTTACTTGAAG 7525

Qy 3481 GCCCTTGCACCGCTGTCGAGCTGCAAAGTCCAGACTGCCACGACTGCTCGGTCGTGTCGGCGGAGAC 3540
          |||||||||||
Db 7526 GCCGGCTGGGCCTGTCGGCGCCTGCGAAGCTCCAGGCGGACTGCGATGGTCGTTATGCGGGACGGGCC 7585

Qy 3541 GGCCCTTGTGTCCTTTATCTTGTGAGAGTCTCCTCGCCCCCCCCCCGGGGGAGACGAGGAGAGTC 3600
          ||||||||||
Db 7586 GACCTTGTCGTTATCTGCTTATTGTTGAAAGCGCCGGGGACGCCGGGACGGCGGGGCGGCC 7645

Qy 3601 TTCACGGAGGGCTATGACTAGGTACTCCTGCCCCCCCCCGGGGGACGAGGGCCGAGGAGAATAC 3660
          |||||||||||||
Db 7646 TTCACGGAGGCTATGACTAGTAGTACTTCTGCCCCCCCCTGGGGAAAACCAGAGAATAC 7705

Qy 3661 GACCTGGAGTTGATAACATCTCCTCCCAAATGTGTTCGGTCGGTCCGCCACCGATGCAATCTGGC 3720
          |||||||||
Db 7706 GACCTGGAGTTGATAACATCATGCCCTCCAAATGTGTTCAGTGCCGACGATGCAATCTGGC 7765

Qy 3721 AAAAGGGGTATACTTACCTCACCCGGTGACCC 3749
          ||||||||||||||||
Db 7766 AAAAGGGGTGTACTACTCACCCGTGACCC 7794
```

```
RESULT 3
US-09-539-601-10
; Sequence 10, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 7989
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I377/NS3-3'/wt
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1181)
; OTHER INFORMATION: hepatitis C virus core-neomycin phosphotransferase
; OTHER INFORMATION: fusion protein
; FEATURE:
```

```
;
;    NAME/KEY: RBS
;    LOCATION: (1190)..(1800)
;    OTHER INFORMATION: internal ribosome entry site from
;    OTHER INFORMATION: encephalomyocarditis virus
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1801)..(7758)
;    OTHER INFORMATION: hepatitis C virus NS3 - 5B
;    FEATURE:
;    NAME/KEY: 3'UTR
;    LOCATION: (7759)..(7989)
;    PUBLICATION INFORMATION:
;    AUTHORS: Lohmann, Volker
;    AUTHORS: Körner, Frank
;    AUTHORS: Koch, Jan-Oliver
;    AUTHORS: Herian, Ulrike
;    AUTHORS: Theilmann, Lorenz
;    AUTHORS: Bartenschlager, Ralf
;    TITLE: Replication of Subgenomic hepatitis c virus RNAs in a
;    TITLE: hepatoma cell line
;    JOURNAL: Science
;    VOLUME: 285
;    PAGES: 110-113
;    DATE: 1999-07-02
;
US-09-539-601-10

Query Match    87.8%;    Score 3291.4;    DB 4;    Length 7989;
Best Local Similarity    92.4%;    Pred. No.: 0;    Mismatches 286;    Indels 0;    Gaps 0;
Matches 3463;    Conservative 0;    Mismatches 286;    Indels 0;    Gaps 0;
```

```
;
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 7
;   LENGTH: 8001
;   TYPE: DNA
;   ORGANISM: Hepatitis C virus
;   FEATURE:
;   NAME/KEY:
;   LOCATION:
;   OTHER INFORMATION: Hepatitis C virus core-neomycin
;   NAME/KEY: 5'UTR
;   LOCATION: (1)..(341)
;   OTHER INFORMATION: phosphotransferase fusion protein
;   FEATURE:
;   NAME/KEY: RBS
;   LOCATION: (342)..(1193)
;   OTHER INFORMATION: construct I389/NS3-3'/wt
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1202)..(1812)
;   OTHER INFORMATION: internal ribosome entry site from
;   OTHER INFORMATION: encephalomyocarditis virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1813)..(7770)
;   OTHER INFORMATION: hepatitis C virus nonstructural proteins NS3-5B
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: (7771)..(8001)
; PUBLICATION INFORMATION:
;   AUTHORS: Lohmann, Volker
;   AUTHORS: Korner, Frank
;   AUTHORS: Koch, Jan-Oliver
;   AUTHORS: Herian, Ulrike
;   AUTHORS: Theilmann, Lorenz
;   AUTHORS: Bartenschlager, Ralf
;   TITLE: Replication of subgenomic hepatitis c virus RNAs in a
;   TITLE: hepatoma cell line
;   JOURNAL: Science
;   VOLUME: 285
;   PAGES: 110-113
;   DATE: 1999-07-02
; US-09-539-601-7
```

```
Query Match      87.8%;  Score 3291.4;  DB 4;  Length 8001;
Best Local Similarity  92.4%;  Pred. No. 0;  Mismatches 286;  Indels 0;  Gaps 0;
Matches 3463;  Conservative 0;  Mismatches 286;  Indels 0;  Gaps 0;
```

```
Qy       1 TGGGAGGAGGGCCTCTTCACAGCCTCACCCACGTGGATGCCCCACTTCCTGTCCCAAACAAAG   60
           ||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db    3769 GGCAGGCGTGGGTCATTGTGGGTAGGAGATCATCTTGTTCCGGGGGCCCGGGCCTATTGTTCCCGAC   420
```

```
Qy    3661 GACCTGGGAGTTGAATGAACATCAAGTGTTCCTCCAAATGTGTCGGTCGGCGGCACGGATGCATCTGGC 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7069 GACTTGGGAGTTGAATGAACATCAAGTGTTCCTCCAAATGTGTCGGTCGGCGGCACGGATGCATCTGGC 7128

Qy    3721 AAAAGGGGTATACTACCCTCACCCGTGACCC 3749
           ||||||||||||||||||||||||||||||||
Db    7129 AAAAGGGGTGTACTATCTCACCCGTGACCC 7157
```

```
Query Match            87.8%;  Score 3291.4;  DB 4;  Length 8637;
Best Local Similarity  92.4%;  Pred. No. 0;  Mismatches 286;  Indels  0;  Gaps  0;
Matches 3463;  Conservative  0;  Mismatches 286;  Indels  0;  Gaps  0;

Qy       1 TGGGAGGGGGGCGTCTTCACGAGGCCGCCTCACCCACGTGGGATGCGCCCCACCTTCGTGTCTCCAAACAAAG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    4045 TGGGAGGGAGGCGGTTCTTTACAGGGCCGTCACCCACGTGGGATGCGCCCCACCTTCGTGTCTCCAAACAAAG 4104

Qy      61 CAGGCAGGAGGACAACTTCCCTACTTGGTGGCGCTACCAGGCTACGTGTGTCGGCTACGGCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4105 CAGGCAGGAGGACAACTTCCCTACTTGGTGGCGCTACCAGGCTACGTGTGTCGGCTACGGCC 4164
```

```
Db    4165 CAGGGTTCCACCTCCACTTCCATTCGTGGGACCAAAAGTGGGAAGTGTTCTCATAGGGTTACGGG 4224
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 CAGGGCCCCACCTCCATTCATTCATGGGGATCAAAAGTGGGAAGTGTTCTCATAGGGTTACT 180

Qy     181 CTGCGCCGGCGCCAACACCCCTTTGCTTGTATAGGTCTGGCGGGTGGCTCCCAAAAACCCTACCC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4285 CTGCACGGGCGGGCCAAGGCCCCTTGCTTGTATAGGTCTGGGCGGGTGGCTCCCAAAAACCGGTTACTACC 4284

Qy     241 ACACACCCCATAACCAAAATTCATCATGGGATGGCAATGGCATGCAACCCCGGGTCGTCACG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4285 ACACACCCCATAACCAAAATACATCATGGCAATGGCAGTCGGCGGTCGTCACG 4344

Qy     301 AGCAACCTGGGTGCGTCGTGTGGGGCGGGGTCCTTGCAGCTCTGGCTCTTGGTATTGCTTGACAACA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4345 AGCAACCTGGGTGCTGGGTGCGTAGCGGGGAGTCGTTGTCAGCAGCTCTGGCTCTTGGTATTGCTTGACAACA 4404

Qy     361 GGCACCGTGGTCATTGTGGTGGTAGGATCATCATTGTCCGGGGGGGGGGCCCCGGC 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4405 GGCACCGTGGTCATTGTGGTGGCAGGATCATCTTTGTCCGGGAAAGGCGCCCATCATTCCCGAC 4464

Qy     421 AGGGAAGTCCTTCTACCCAGAGAGTTCGATGAGGAGATGGAGGAGTGGGCCGCAACCTTGCGGGG 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4465 AGGGAAGTCCTTTACGGGAGTTCGATGAGATGGCGGAGAAGGAGTGGGCCGGGCCGCAGCACCTCCTTAC 4524

Qy     481 ATCGAGCACGGGAATGCAGCTCGTCGGCCCAGCACGTTCAAGCAAAAAGGCGTCGGGGTTGCAG 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4525 ATCGAACAGGGAAATGCAGCTTCGCGGACAATTCAAAACAGAAAAGGCAATCGGCGTTGGCAA 4584

Qy     541 ACAGCCACCAAGCAAGCAGCAGAGGGCCGCTGCTCCCGGTGTGGTGGAGTCCAAGTGGGCGACCCTT 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    4585 ACAGCCACCAAGCAAGCAGAGGGCCGCTGCTCTCCCGGTGGGTGGGAAATCCAAGTGGCGGACACCTC 4644

Qy     601 GAGAGCTTCTTGGGGCGGAAACACATGTGGAACATTCATCAGCGGCGGGGGATACAGTTACTTAGGCAGGC 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4645 GAAGCCTTCTCTGGGCGGAAAGCAGCATATGTGGAAATTTCACACAGCGGCCGGATACAGTAATTAGGCAGGC 4704

Qy     661 TTGTCCACTCTTGCCTGGGGAATCCCCGGCGGATTGCATCACTGAGGGGTTCACAGCCTTCGTGTC 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4705 TTGTCCAACTCTGCCTGGGCAACCCCCGGATAGCATCAACTGATGGGGTTCACAGGCCTCTATC 4764

Qy     721 ACTAGCCCGCTCACCCACCCAAATCTCTACCCTCCTGCTTAACATCCCTGGTTAACATCCTG 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4765 ACCAGCCCGCGTCACCCCAACCAAATCCCTGGTTTAACATCCCTGTTTAACATCCTGTTTAACATCCTGGCC 4824

Qy     781 GCCCCAACTCGCTCCCCCCCAGTGTCGCTTCAGCTTTCGTTAGGGCGGCGGCCAGCATTGCTCGGTGCG 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4825 GCCCAACTTGCTCGTCCTCCCCAGGGTTCTTCCTCTCCGTTTCGGTTAGGGCGGCGGCCATCGGCTGCGGAGCG 4884

Qy     841 GCTGTGTTGGCAGCATAGGGCCTTGGGGAAGGTGGCTTGGGTTGGTGGAACATCTGGCCGCCCTATGGAGCA 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4885 GCTGTTGGCAGCATAGGCCCTTGGGAAGGTGCTTGGGTTGTTGGAATATTTGGCAGGTTATGGAGCA 4944

Qy     901 GGAGTTGGCAGGCGCGCGTCGTGTCGGTCCTTTAAGGTCATCATGAGCGCGGGAAATGCCCTCCACCGAG 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4945 GGGGTTGGCAGGCGCGCGTCGTGTCGGTCCTTTAAGGTGTCATGAGCGCGGAGATGCCCTCCACCGAG 5004

Qy     961 GACCTGGTTAACTTACTCCCGACTCCGTCCTCCCTGCCATCCTCTCTCCTGGGGTGCCCTGGCCCTGGTTCG 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5005 GACCTGGTTAACCTACTCCCTGCTATCCTCTCGACGCCTTTAAGGTCATGAGCGCGGAGATGCCCCACCGAG 5064

Qy    1021 TGCGCAGCGAGTACTGCGTCGGCGCACGTGGGTCCAGGGGAGGGGGCTGTGCGAGTGGATGAAC 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5065 TGCGCAGCGAGATACTGCGTCGGGCCGCCAGCGTGGCCGTGGGGCCTTGTGCAGTGGATGAAC 5124

Qy    1081 CGGCTGATAGGGTTCGGCTCGGCCTCGGGTAACCACGTGTTTTCCCCCACGACTATGTTGCCAGAG 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5125 CGGCTGATTAGCGTTCGGCTTCGCGGGGGGTAACCACGTGTTTTCCCCACGACTATGTTGCTGAG 5184

Qy    1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTTCTCCGACACTTTACTATCACCCGACTCGTTG 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5185 AGCGACGCCGCAGCACGTGTCACTCAGATCTCTCCGACACTTTACTATCACCCGACTCGTTG 5245
```

```
Db   5185 AGCGACGCTGCAGCACGTGTCACTCAGATCCTCTCTAGTCTTACCATCACTCAGCTGCTG 5244

Qy   1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCACGCCCTGCTCCGGCTCGTGGCTA 1260
          |||||||| ||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db   5245 AAGAGGCTTCACCAGTGGATCAACGAGGACTGCTCCACGCCATGCTCCGGCTCGTGGCTA 5304

Qy   1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
          || ||||||||||| ||||||||||| || ||| |||| ||||||||||||||||||||||
Db   5305 AGAGATGTTTGGGATTGGATATGCACGGTGTTGACTGATTTCAAGACCTGGCTCCAGTCC 5364

Qy   1321 AAGCTCCTGCCGCGATTACCGGGAGTCCCCTTTTTCTCATGCCAACGTGGGTACAAGGGG 1380
          ||||||||||||||||| |||||||||||||| |||||||| ||||||||||||||||||
Db   5365 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 5424

Qy   1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440
          ||||||||||| |||||||||||||| ||||||||| ||||||||||||||||||||||||
Db   5425 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 5484

Qy   1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
          ||||| ||||||||||||||||||||||| ||||||| |||||||||||||| |||||||
Db   5485 CATGTGAAAAACGGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5544

Qy   1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||| ||||||||
Db   5545 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 5604

Qy   1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5605 TATTCTAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 5664

Qy   1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACGTAAAATGCCCGTGCCAGGTTCCA 1680
          |||||||||||||||||| ||||||||||||||||||||||| |||||||| ||||||||
Db   5665 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5724

Qy   1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTGC 1740
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db   5725 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGCTCCAGCGTGC 5784

Qy   1741 AAACCTCTCCTACGGGAGGAGGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTTGGG 1800
          ||||| ||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||
Db   5785 AAACCCCTCCTACGGGAGGAGGTCACATTCCTGGTCGGGCTCAATCAATACCTGGTTGGG 5844

Qy   1801 TCGCAGCTCCCATGCGAGCCCGAACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 1860
          || ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   5845 TCACAGCTCCCATGCGAGCCCGAACCGGACGTAGCAGTGCTCACTTCCATGCTCACCGAC 5904

Qy   1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920
          ||||||||||| || || |||||||||||||| ||||||||||||||| ||||||||||||
Db   5905 CCCTCCCACATTACGGCGGAGACGGCTAAGCGTAGGCTGGCCAGGGGATCTCCCCCCTCC 5964

Qy   1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980
          ||||||||||| |||||||||||| ||||||||||||||||| |||||| |||| || ||||
Db   5965 TTGGCCAGCTCATCAGCTAGCCAGCTGTCTGCGCCTTCCTTGAAGGCAACATGCACTACC 6024

Qy   1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040
          |   |||||| ||| ||||||||||||||||||||||||||||||||||||||| ||||||
Db   6025 CGTCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATG 6084

Qy   2041 GGCGGGGACATTACCCGCGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
          |||||| |||| |||||||||||||||||| || |||||||||||  |||||||||||||
Db   6085 GGCGGGAACATCACCCGCGTGGAGTCAGAAAATAAGGTAGTAATTTTGGACTCTTTCGAG 6144

Qy   2101 CCGCTCCGAGCGGAGGAGGATGAGCGGGAAGTGTCCGTCCCGGCGGAGATCCTGCGGAAA 2160
          ||||||| |||||||||||||||| ||||||| ||||| ||||||||||||||||||||
Db   6145 CCGCTCCAAGCGGAGGAGGATGAGAGGGAAGTATCCGTTCCGGCGGAGATCCTGCGGAGG 6204

Qy   2161 TCCAAGAAATTCCCACCAGCGATGCCCGCATGGGCACGCCCGGATTACAACCCTCCGCTG 2220
          |||| ||||||||| | ||||||||||  ||||||||||||||||||||||||||||| |||
Db   6205 TCCAGGAAATTCCCTCGAGCGATGCCCATATGGGCACGCCCGGATTACAACCCTCCACTG 6264

Qy   2221 CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGTGCCCACTGCCA 2280
           | |||||||||||||| ||||||||||||||||||||||||||| |||||| ||| ||||
Db   6265 TTAGAGTCCTGGAAGGACCCGGACTACGTCCCTCCAGTGGTACACGGGTGTCCATTGCCG 6324

Qy   2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340
          ||| | ||| ||||||| |||||||||||||||||||||||||||| |||||| ||| |||||
Db   6325 CCTGCCAAGGCCCCTCCGATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGTCAGAA 6384

Qy   2341 TCCACCGTGTCTTCTGCCCTGGCGGAGCTTGCCACAAAGGCTTTCGGTAGCTCCGAACCG 2400
          || ||||||||||||||| |||||||||| ||||||||| | ||||| ||||||||| ||
Db   6385 TCTACCGTGTCTTCTGCCTTGGCGGAGCTCGCCACAAAGACCTTCGGCAGCTCCGAATCG 6444

Qy   2401 TCGGCCGTCGACAGCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACGACGGCGGA 2460
          |||||||||||||||||||||||||| ||| ||||||||||| ||||||||||||||||
Db   6445 TCGGCCGTCGACAGCGGCACGGCAACGGCCTCTCCTGACCAGCCCTCCGACGACGGCGAC 6504

Qy   2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAC 2520
          || ||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   6505 GCGGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAT 6564

Qy   2521 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGTGAGGAGGCCGGTGAGGACGTCGTC 2580
          |||||||||||||||||||||||||||||||| || ||||||||  ||||||||||||||||
Db   6565 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTAAGCGAGGAGGCTAGTGAGGACGTCGTC 6624

Qy   2581 TGCTGCTCGATGTCCTACACATGGACAGGCGCTCTGATCACGCCATGCGCTGCGGAGGAA 2640
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   6625 TGCTGCTCGATGTCCTACACATGGACAGGCGCCCTGATCACGCCATGCGCTGCGGAGGAA 6684

Qy   2641 AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTAC 2700
          | ||||||||||||||| ||  |||||||||||||||||| ||||||||||||| |||||||
Db   6685 ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCCGTCACCACAACTTGGTCTAT 6744

Qy   2701 GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAA 2760
          ||||| ||||| |||||||||||||| |||||||||||||||||||||||||||||||||
Db   6745 GCTACAACATCTCGCAGCGCAAGCCTGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAG 6804

Qy   2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 2820
           |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db   6805 GTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 6864

Qy   2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAAA 2880
          |||||||| |||||||| || ||||||||||| ||||||||||||||||||||||||||| |
Db   6865 AAGGCTAAACTTCTATCCGTGGAGGAAGCCTGTAAGCTGACGCCCCCACATTCGGCCAGA 6924

Qy   2881 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   6925 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCGTTAACCAC 6984

Qy   2941 ATCCGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000
          ||||||||||||||| |||||||| |||||||||||||| |||||||||||||||||||||
Db   6985 ATCCGCTCCGTGTGGAAGGACTTGCTGGAAGACACTGAGACACCAATTGACACCACCATC 7044

Qy   3001 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGAGGAGGCCGCAAGCCAGCTCGC 3060
          |||||||||||||||||||||||||||||||||||||  || ||||||||||||||||||
Db   7045 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAGGGGGGCCGCAAGCCAGCTCGC 7104

Qy   3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGTGTGTGCGAGAAAATGGCCCTCTATGACGTG 3120
          |||||||| ||||||| |||||||| ||||||||||||||||||||||||| || || |||
Db   7105 CTTATCGTATTCCCAGATTTGGGGGTTCGTGTGTGCGAGAAAATGGCCCTTTACGATGTG 7164

Qy   3121 GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTACGGATTCCAGTATTCTCCTGGA 3180
          |||||||||||||||||||| ||||||||||| || ||||||||||| || |||||||||
Db   7165 GTCTCCACCCTCCCTCAGGCCGTGATGGGCTCTTCATACGGATTCCAATACTCTCCTGGA 7224

Qy   3181 CAGCGGGTCGAGTTCCTGGTGAACGCCTGGAAATCAAAGAAGACCCCTATGGGCTTTGCA 3240
          ||||||||||||||||||||||| |||||||||| | |||||   ||||||||||||| |||
Db   7225 CAGCGGGTCGAGTTCCTGGTGAATGCCTGGAAAGCGAAGAAATGCCCTATGGGCTTCGCA 7284

Qy   3241 TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300
          |||||||||||||||||||||||||| ||||||||||||||||||||||||| |||||||||
Db   7285 TATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 7344

Qy   3301 ATTTATCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 3360
          || || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7345 ATCTACCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 7404
```

```
RESULT 6
US-10-029-907-24
Sequence 24, Application US/10029907
Patent No. 6706874
GENERAL INFORMATION:
  APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
  TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
  TITLE OF INVENTION: HEPATITIS C VIRUS
  FILE REFERENCE: 13/083
  CURRENT APPLICATION NUMBER: US/10/029,907
  CURRENT FILING DATE: 2001-12-21
  PRIOR APPLICATION NUMBER: 60/257,857
  PRIOR FILING DATE: 2000-12-22
  NUMBER OF SEQ ID NOS: 25
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 24
  LENGTH: 8638
  TYPE: DNA
  ORGANISM: HCV
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (1802)...(8407)
US-10-029-907-24

Query Match    87.8%;  Score 3291.4;  DB 4;  Length 8638;
Best Local Similarity   92.4%;  Pred. No. 0;  Mismatches 286;  Indels 0;  Gaps 0;
Matches 3463;  Conservative 0;  Mismatches 286;  Indels 0;  Gaps 0;
```

```
Qy      1  TGGGAGGGGGCGTCTTCACAGGCCTCACCCACGTGATGAGCCCACTTCCTGTTCCCAAACAAAG    60
Db   4046  TGGGAGAGCGGTCTTTTACAGGGCCTCACCCACCATTTCTTGTCCCAGACTAAG        4105

Qy     61  CAGGGAGGAGAGACAACTTCCCCTACCTGGTGTGGGCGTACCAGCAGCCTACTGTGTCGGCCTAGGGCG   120
Db   4106  CAGGGCAGGAGACAACTTCCCCTACCTGGTAGCATACCAGCAGCCTGTGTCGGCCTAGGGGCT   4165

Qy    121  CAGGGCCCACCTCCATCATGGGATCAAAATGTGGAAGTGTCTCATCATCACGGGACTA   180
Db   4166  CAGGGCTCCACCTCCATCGGGGACCAAAAGTGGAAGTGTCTCATCACGGGCTAAAAGCCTACG   4225

Qy    181  CTTGGCGCGGGCCAACACCCTTGCTTGTATATAGGCTGGGAGCCGGTTCAAAAGCTCACCCTC   240
Db   4226  CTGCACGGGCCGGCCAACCGCCCTGCTGTATAGGCTGGGAGCCGGTTCAAAGGAGGTTACTACC   4285

Qy    241  ACACACCCCCATAACCAAAATTCATCAATGGCATGCATGTCAGCCGGACCTGGACCTCGGTCGGTCACG   300
Db   4286  ACACACCCCATAACCAAATACATCAATGGCATGTCAGCCGGACCTGGACGTCGTCACG   4345

Qy    301  AGCACCTGGGTGTCTGCTGGTGCGTAATACCCTCACAACATGTTACTTGAG   360
Db   4346  AGCACCTGGGTGCTGGTGTCGTAAGAGTCCTAGGGAGGTCTGCGAGAACA   4405

Qy    361  GGCAGCGTTGGTCATTGTGGGGTAGGATCATCATCTTGTTCCGGGGGGGCTATTGTTCCCGGAC   420
Db   4406  GGCACGGTGGTCATTGTGGGCAGGGATCATCTTGTCCGGAAAGCCGGCCATCATTCCGAC   4465

Qy    421  AGGGAAGTTCCTTCTACCGAGGAGTTCGATGAGATGAGATGGAAGAGTGCGGTCGGTCGCACCTTCCCTTAC   480
Db   4466  AGGGAAGGTCCTTTACCAGGAGTTCGATGAGATGGAAGAGTGCGCTCGTTCGCACCTTAC   4525

Qy    481  ATCGAGCAGGAAATGCAGTTCAGCTGGCGGGGAGTTGAAGCAAAAAAGCGTCCGGTTGCGCAA   540
Db   4526  ATCGAACAGGGAAATGCAGTCAGCTGGCCGCGGGAGTTGCAA   4585

Qy    541  ACAGCCACCAAGACAAGACGGAGGGCCGCTGCCCGTGGGTGGAAGTGGGCGAGCCCTT   600
Db   4586  ACAGCGACCACCAAGCAGAGGAGCTGTCGCCCGTGGTGAATCCAAGTGGCGGGACCCTC   4645

Qy    601  GAGACCTTCTGGGCGAAAACACATGTGGGAACTTCATCAGCAGGATACAGTACTTAGGAGGC   660
Db   4646  GAAGCCTTCTGGGCCGGAAAGCATTATGTGGAATTCATCAGCAGGAATATTTAGCAGGC   4705

Qy    661  TTGTCCACTCTTGCCTTGGGCGGATTACCCCGGATTTGCATCACTGAGTTCAGCCCTCTGTC   720
Db   4706  TTGTCCACTCTTGCTTGGGCCGCCGTTAGCACATGCATTCACAGCCCTATC   4765

Qy    721  ACTAGCCCGGCTTCACCAGTTCCGCCAACCCCGGGATAGCATCTCTGTTTAACATCCTCGTTTGCTTCCC   780
Db   4766  ACCAGCCGGCTTCACCAGCCAACATACCCTCGTTTAACATCCTCGTTTGGCC   4825

Qy    781  GCCCAACTCCGCTCCCCCCCCAGTGCTGCTTCAGCTTTCGTAGGCGCCGCAATTGCTGGTCG   840
Db   4826  GCCCAACTTGCTCCTCCCCAGGCCGCTGCTTCTGCTAGGCGCCATGGCATTGGTCG   4885

Qy    841  GCTCGTTGGCAGCATAGGCCCTTGGGGAAGGTGCTTGTTGTTGGACAATCTTGGCGCCGGTATGGAGCA   900
Db   4886  GCTCGTTGGCAGCATAGGCCTTTGGGAGGTGCTTTGTGGATATTTGGCAGGTTATGGAGCA   4945

Qy    901  GGAGTGGCAGGGCGCTCGTGTGGCCTTAAGGTCATGAGCGGCGAAATGCCCTCCCACCGAG   960
Db   4946  GGGGTGGCAGGCGATACTCGGGCCTTGGGGCCTTCGAGCGGCAGAATGCCCTCCACCGAG   5005

Qy    961  GACCTGGTTAACTTACTCCCATCCTCTCTGTTGGTGCCCTGGTTGTCGTCGTG   1020
Db   5006  GACCTGGTTAACCTACTCCCCATCCTATCCTTGGTGCCCTGGTTGTCGTCGTG   5065

Qy   1021  TGCCGCAGCGATACTCGGGGTCGGGGTCCAGCGGAGGGGGCTCGCACTGGGATGAAC   1080
Db   5066  TGCCGCAGCGATACTCGGGGTCGGGGTCCAGCGGAGGGGGCTGTGCACTGGGATGAAC   5125

Qy   1081  CGGCTGATAGGCGTTCGCCTCGGGGGGGTAACCANTGTTCCCCAGCCCACTATGTGCCCAGAG   1140
Db   5126  CGGCTGATAGGCGTTCGCCTCGGGGGGGTAACCACGTGTTCCCCAGCACTATGTGCCGAG   5185

Qy   1141  AGCGACGGCGGCAGCAGTGTCACTCAGATCCTTCTCGACTTTACATCACCCAAACTGTTG   1200
Db   5186  AGCGACGGCTGCAGCAGTGTCACTCAGATCTCCTCTCCGGACTCTTTACTATCACACCAAACTGTTG   5245

Qy   1201  AAGAGGCTCCACCAGTGGGGATTAACGGACTGCTCCACGCGCCCGTTGCGTGGCTA   1260
Db   5246  AAGAGGCTTCACCAGTGGGATCAACGAGGACTGCTCCAGCGCGCCATGCTCCGTGTGGCTA   5305

Qy   1261  AGGGATGTTTGGGACTGGGATATGCACAGTTTGGCTGGACTTCAAGAACCTGGGCTCC   1320
Db   5306  AGAGATGTTTGGGATTGGGATATGCACAGTTGTTGACTGAATTTCAAGAACCTGGGCTCC   5365

Qy   1321  AAGCTCCTCGCCGCGATTACCGGGGAGTCCCCCTTTTCTCATCCCAAGTGGGGTACAAGGGG   1380
```

```
              ||||||||||||||||| |||||||||||||| |||||||| |||||||||||||||||
Db       5366 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 5425

Qy       1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440
              ||||||||||| |||||||||||||| ||||||||| |||||||||||||||||||||||
Db       5426 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 5485

Qy       1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
              ||||| ||||||||||||||||||||||| ||||||| |||||||||||||| |||||||
Db       5486 CATGTGAAAAACGGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5545

Qy       1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560
              |||||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||
Db       5546 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 5605

Qy       1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5606 TATTCTAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 5665

Qy       1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACGTAAAATGCCCGTGCCAGGTTCCA 1680
              |||||||||||||||||| |||||||||||||||||||||| |||||||| ||||||||
Db       5666 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5725

Qy       1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTGC 1740
              |||||||||||||||||||||||||||||||||||| |||||||||||||||| ||||||
Db       5726 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGCTCCAGCGTGC 5785

Qy       1741 AAACCTCTCCTACGGGAGGAGGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTTGGG 1800
              ||||| ||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db       5786 AAACCCCTCCTACGGGAGGAGGTCACATTCCTGGTCGGGCTCAATCAATACCTGGTTGGG 5845

Qy       1801 TCGCAGCTCCCATGCGAGCCCGAACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 1860
              || |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       5846 TCACAGCTCCCATGCGAGCCCGAACCGGACGTAGCAGTGCTCACTTCCATGCTCACCGAC 5905

Qy       1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920
              ||||||||||| || || |||||||||||||| |||||||||||||| ||||||||||||
Db       5906 CCCTCCCACATTACGGCGGAGACGGCTAAGCGTAGGCTGGCCAGGGGATCTCCCCCCTCC 5965

Qy       1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980
              ||||||||||| |||||||||||| ||||||||||||||| |||||| |||| || ||||
Db       5966 TTGGCCAGCTCATCAGCTAGCCAGCTGTCTGCGCCTTCCTTGAAGGCAACATGCACTACC 6025

Qy       1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040
              |   |||||| ||| |||||||||||||||||||||||||||||||||||||| ||||||
Db       6026 CGTCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATG 6085

Qy       2041 GGCGGGGACATTACCCGCGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
              |||||| |||| ||||||||||||||||| || |||||||||||  |||||||||||||
Db       6086 GGCGGGAACATCACCCGCGTGGAGTCAGAAAATAAGGTAGTAATTTTGGACTCTTTCGAG 6145

Qy       2101 CCGCTCCGAGCGGAGGAGGATGAGCGGGAAGTGTCCGTCCCGGCGGAGATCCTGCGGAAA 2160
              ||||||| |||||||||||||||| ||||||| ||||| |||||||||||||||||||
Db       6146 CCGCTCCAAGCGGAGGAGGATGAGAGGGAAGTATCCGTTCCGGCGGAGATCCTGCGGAGG 6205

Qy       2161 TCCAAGAAATTCCCACCAGCGATGCCCGCATGGGCACGCCCGGATTACAACCCTCCGCTG 2220
              |||| ||||||||| | ||||||||||  ||||||||||||||||||||||||||| |||
Db       6206 TCCAGGAAATTCCCTCGAGCGATGCCCATATGGGCACGCCCGGATTACAACCCTCCACTG 6265

Qy       2221 CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGTGCCCACTGCCA 2280
               | ||||||||||||| ||||||||||||||||||||||||||| ||||| ||| ||||
Db       6266 TTAGAGTCCTGGAAGGACCCGGACTACGTCCCTCCAGTGGTACACGGGTGTCCATTGCCG 6325

Qy       2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340
              ||| | ||| ||||||| |||||||||||||||||||||||||| ||||| ||| |||||
Db       6326 CCTGCCAAGGCCCCTCCGATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGTCAGAA 6385

Qy       2341 TCCACCGTGTCTTCTGCCCTGGCGGAGCTTGCCACAAAGGCTTTCGGTAGCTCCGAACCG 2400
              || ||||||||||||||| |||||||||| ||||||||| | ||||| ||||||||| ||
Db       6386 TCTACCGTGTCTTCTGCCTTGGCGGAGCTCGCCACAAAGACCTTCGGCAGCTCCGAATCG 6445

Qy       2401 TCGGCCGTCGACAGCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACGACGGCGGA 2460
              |||||||||||||||||||||||||| ||| |||||||||| ||||||||||||||||
Db       6446 TCGGCCGTCGACAGCGGCACGGCAACGGCCTCTCCTGACCAGCCCTCCGACGACGGCGAC 6505

Qy       2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAC 2520
              || ||||| |||||||||||||| |||||||||||||||||||||||||||||||||||
Db       6506 GCGGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAT 6565

Qy       2521 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGTGAGGAGGCCGGTGAGGACGTCGTC 2580
              |||||||||||||||||||||||||||||||| || ||||||||  ||||||||||||||
Db       6566 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTAAGCGAGGAGGCTAGTGAGGACGTCGTC 6625

Qy       2581 TGCTGCTCGATGTCCTACACATGGACAGGCGCTCTGATCACGCCATGCGCTGCGGAGGAA 2640
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       6626 TGCTGCTCGATGTCCTACACATGGACAGGCGCCCTGATCACGCCATGCGCTGCGGAGGAA 6685

Qy       2641 AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTAC 2700
              | ||||||||||||||| ||  |||||||||||||||| |||||||||||| |||||||
Db       6686 ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCCGTCACCACAACTTGGTCTAT 6745

Qy       2701 GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAA 2760
              ||||| ||||| ||||||||||||| |||||||||||||||||||||||||||||||||
Db       6746 GCTACAACATCTCGCAGCGCAAGCCTGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAG 6805

Qy       2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 2820
               |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db       6806 GTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 6865

Qy       2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAAA 2880
              |||||||| |||||||| || ||||||||||| ||||||||||||||||||||||||| |
Db       6866 AAGGCTAAACTTCTATCCGTGGAGGAAGCCTGTAAGCTGACGCCCCCACATTCGGCCAGA 6925

Qy       2881 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       6926 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCGTTAACCAC 6985

Qy       2941 ATCCGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000
              ||||||||||||||| |||||||| ||||||||||||| |||||||||||||||||||||
Db       6986 ATCCGCTCCGTGTGGAAGGACTTGCTGGAAGACACTGAGACACCAATTGACACCACCATC 7045

Qy       3001 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGAGGAGGCCGCAAGCCAGCTCGC 3060
              |||||||||||||||||||||||||||||||||||||  || ||||||||||||||||||
Db       7046 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAGGGGGGCCGCAAGCCAGCTCGC 7105

Qy       3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGTGTGTGCGAGAAAATGGCCCTCTATGACGTG 3120
              |||||||| |||||||| |||||||| ||||||||||||||||||||||| || || |||
Db       7106 CTTATCGTATTCCCAGATTTGGGGGTTCGTGTGTGCGAGAAAATGGCCCTTTACGATGTG 7165

Qy       3121 GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTACGGATTCCAGTATTCTCCTGGA 3180
              |||||||||||||||||||| ||||||||||| || ||||||||||| || |||||||||
Db       7166 GTCTCCACCCTCCCTCAGGCCGTGATGGGCTCTTCATACGGATTCCAATACTCTCCTGGA 7225

Qy       3181 CAGCGGGTCGAGTTCCTGGTGAACGCCTGGAAATCAAAGAAGACCCCTATGGGCTTTGCA 3240
              ||||||||||||||||||||||| ||||||||| | |||||   |||||||||||| |||
Db       7226 CAGCGGGTCGAGTTCCTGGTGAATGCCTGGAAAGCGAAGAAATGCCCTATGGGCTTCGCA 7285

Qy       3241 TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300
              |||||||||||||||||||||||||| ||||||||||||||||||||||| |||||||||
Db       7286 TATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 7345

Qy       3301 ATTTATCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 3360
              || || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7346 ATCTACCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 7405

Qy       3361 CGGCTTTATATCGGGGGTCCCCTGACTAATTCAAAAGGGCAGAACTGCGGCTATCGCCGG 3420
              |||||||| |||||||| |||||||||||||| |||||||||||||||||||||||||||
Db       7406 CGGCTTTACATCGGGGGCCCCCTGACTAATTCTAAAGGGCAGAACTGCGGCTATCGCCGG 7465

Qy       3421 TGCCGCGCGAGCGGCGTGCTGACGACTAGCTGCGGTAATACCCTCACATGTTACTTGAAG 3480
              |||||||||||||| || |||||||| |||||||||||||||||||||||||||||||||
Db       7466 TGCCGCGCGAGCGGTGTACTGACGACCAGCTGCGGTAATACCCTCACATGTTACTTGAAG 7525

Qy       3481 GCCTCTGCAGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTCGTGTGCGGAGAC 3540
              ||| |||| |||||||||||||| |||||||||||||||||||||||||| |||||||||
Db       7526 GCCGCTGCGGCCTGTCGAGCTGCGAAGCTCCAGGACTGCACGATGCTCGTATGCGGAGAC 7585
```

RESULT 7
US-10-029-907-1
; Sequence 1, Application US/10029907
; Patent No. 6706874
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; TITLE OF INVENTION: HEPATITIS C VIRUS
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 8639
; TYPE: DNA
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1803)...(8408)
US-10-029-907-1

Query Match   87.8%;   Score 3291.4;   DB 4;   Length 8639;
Best Local Similarity   92.4%;   Pred. No. 0;
Matches 3463; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

```
Qy      1 TGGGAGGGGGGCTCTTCACAGGCCCTCACCCACGTGGATGCCCACTTCCTGTCCCAAACAAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4047 TGGGAGGAGGGGTCTTTACAGGCCCTCACCCACATAGACGCCCATTTCTTGTCTCCAGACTAAG 4106

Qy     61 CAGGCAGGAGACAAACTTCCCCCTACCTACCTGGTGGGGGTACCAGGCTACTCGTGTTGTGCGCGGTTGCGGTGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4107 CAGGCAGGAGAGACAACTTCCCCTACCTGGTGTAGCATACACAGGCTACGGTGTGCGCCAGGGGCT 4166

Qy    121 CAGGCCCCACCTCCATCATGGGGGATCAAAATGTGGAAAGTGTCTCATACGGGCTAAAGCCTACT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4167 CAGGCTCCACTCCACTGTGGGACCCAAAATGGAAGTGTTCTCATACGGGGCTAAAGCCTACG 4226

Qy    181 CTGCGGCGGGCCACACCCTTGCTTGTATAGGCTGGGAGCCGTTCAAAAACGAGGTCACCCCTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4227 CTGCACGGGGCCAAGCCCCTGCTGTATAGGCACCGGTTCAAAAACGAGGTTACTACC 4286

Qy    241 ACACACCCCATAACCAAATTCATCATGGCATGGCAATGTCAGCCGACCTGGAAGGTCGCTCACG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4287 ACACACCCCATAACCAAAATACATCATGGCCAATGTCAGCCGACCTGGAGGGTCGTCACG 4346

Qy    301 AGCACCTGGGTGTCTGGTTGGTCGGGGGTGCTCAGCTCTTGCAGCTGTATTGGTTGACAAACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4347 AGCACCTGGGTGCTGGTTGGGGGGGACTCCTTAGCAGCTCTGGCCTATTTGCCTGGACACCGGA 4406

Qy    361 GGCACGGCGTGGTCATGTCATGTGGGGCAGGAGGATTCATCTCTTGTTCGGGGAAAGCCGGCCCGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4407 GGCAGCGGTGGTCTCATTGGGGTAGGATTCATCTCTTGTTCCCGGATCATCATTCCCGGAC 4466

Qy    421 AGGGAAGTTCCTCTTCACCAGAGAGTTCGATGGAGTTGAGATGGGGGAGAGGAGAGTGCGGCGTCGCCACCTTAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4467 AGGGAAAGTTCCTTTACCCGGAGGAGTTCGATGAGATGGGAAAGAGTGCGGCCTCACACCTCCCTTAC 4526

Qy    481 ATCCAGACGCAGGGAAATGCAGCTCGCCGGCAGCTCCGCCGGAGCGCAGCAGTTCAAGCAGCAAAAAGGCTTCGGGTTGCTGCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4527 ATCGAACAGGGAAATGCAGTCTGCCCGAACAATTCAAACAGAAATCAGAATAC 4586

Qy    541 ACAGCCACCAAAGCAAGCAGGAGGAGGCCGGCTGCTCCCGCTGTCCCGGTCCAAGTGGGGCCCCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4587 ACAGCCACCAAAGCAAGCGAGGGGCCTGCTCCCCGGGTGGGAATCCAAGTGGCGGACCCTC 4646

Qy    601 GAGACCTTCTGGGGGCCGAAACACATGTGGACTTCATCAGCAGGTACTACTTAGCAGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4647 GAAGCCTTCTGGGGGAAGCATATGTGGAATTTCATCAGCGGGGATACAAATATTTAGCAGGC 4706

Qy    661 TTGTCCCACTCTGCCTCTGGGAATCCCGGCATGTGCATCACTACTTGCAGGCCTTCTGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4707 TTGTCCACTCTGCCTCTGGCAACCCCGGCATAGCATCACTGATGGGCATTCACAGCCTTCTATC 4766

Qy    721 ACTAGCCCGGTCACCCAAAATCTACCCTCCTTAACATCCTGGGGGGGATGGGGTAGCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4767 ACCAGCCCGGTCACCCAAAATCTACCCTCCTGTTAACATCCTGGGGATGGGGTAGCC 4826

Qy    781 GCCCAAACTCGCTCCCCCCAGTGCTTCCTTCAGCTTCGTTAGGCGCCCGGCCATTGCTGCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4827 GCCCAAACTTGCTCCCTCCCAACGCGTGCCTTTCTTGTTGTGGACATCTTGGCGCCCGGCA 4886

Qy    841 GCTGTTGGGCAGCATAGGCCTTGGGAAGGTGCTTGTTGGCCTTTGGCGCCGCGCCATCGGCTGGAGCG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4887 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTTGTGTCGGGATATTTGGCAGGTTATGAGCA 4946

Qy    901 GGAGTGGCCAGGCGGCGCCTCGTGGCCTTTAAGGTCATCGTAGTGAGCGGAAATGCCCTCACCGAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4947 GGGGGTGGCAGGCGCGGCCTCGTTGGCCTTTGAAGGTCATGAGCGGGTGCAGTGGGAGGAAC 5006

Qy    961 GACCTGGTTAACTTACTCCCTGCCATCCTCTCCTGGGCCAGTGGTCCCCTGGTCGTCGTGGGGGTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5007 GACCTGGTTAACCTACTCCCTGCTATCCCCTGGGCCCAGTAGCCACTAATGTCGTCGTG 5066

Qy   1021 TGGCGCAGGCGGATTACGTCCGGTCGGGCACGTGTCGGGTTCAGGAGGGAGGGGGGGTCGCAGTGGGGAATGAAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5067 TGCGGACGGACGATACTCGGTCGGCACCTGGGAGGGAGGGGGGGCTGTGCAGTGGGATGGAAC 5126

Qy   1081 CGGGCTGATAGGCGGTTGGCGCCCTGCCTGCGGGGGGTAACCCATGGTTTCCCCGAACCACGAGTGCCTGAG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5127 CGGGCTGATAGGCGGTTCGCTTCGGCACTGGTCGGCACCTGGGCCAGGGGAGGGGCTGTGCAGTGGGATGGAAC 5186

Qy   1141 AGGCGACGGCGCCAGCAGCGTCACTCGGGGGGGTAAGCCTGCCCCGGGGTACAAAGGGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5187 AGGCGACGGCGCCAGCAGCAACGTGTCACTCGGGGCTACTGCCTTCTACCATCAGTGTCTG 5246

Qy   1201 AAGAGGCTTCCACCAGTGGATTAACGGAGAGTTGCTCCAGCGCCGCCCTTGCTCCTCTGGGGGCTA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5247 AAGAGGCTTCACCAGTGGGAATCCCAGGGAAGGATGCTCGCCCCAGGCCCATGCTCCGGGTCGGTA 5306

Qy   1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGGGTGACTTCAAGACCTGGCTCCAGTCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5307 AGAGATGTTTGGGACTGGATATGCACAGTTTTGGGGTTGACTTCAAGACCTGGCTCCAGTCC 5366

Qy   1321 AAGCTCCTGCCGCGGGACTCCCCTTTTTCTCATGCGCAACGTGGGGTACAAAGGGG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5367 AAGCTCCTGCCGGGGGAGTGCCCCTTTCTTCTCATGTCAACGTGGGGACAAAGGGGA 1440

Qy   1381 GTCTCGGCGGGGAGACGGCGGCATCATGCAGGACCACCTCCATCGTGATGGGGCACACCGGA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5427 GTCTCGGCGGGGACGGCGGCATCGAAACCCACCCTCCAGTGGAGCACAGATGCACCGGA 5486

Qy   1441 CATGTCAAAAACGGTTCCATGAGGAGGATCGTTGGGGGCCTAAGACCTGTAGTGAACACGTGGCAT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5487 CATGTTGAAAAACGGTTCCATGAGGGGATTCGTGTGGGCCCTAGGACCTGTAGTAACACGTGGCAT 5546
```

```
Qy   3541 GGCCTTGTCGTCGTTATCTGTCAGAGAGCGGGGCGGGGAAGGAGGACGGGGGGGGCCGAGCCTACGGAGTC 3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7586 GACCTTGTCGTTATCTGTGAAAAGGCGGGGGGACCCAAGAGGACGGAGGCCTACGGGGCC 7645

Qy   3601 TTCACGCGAGGCTAGCTATTGACTAGGTACTCTTCGCCCCCCCCCCCGGGGGGGCCCAAGCCCCAAGAATAC 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7646 TTCACGCGAGGCTATGACTAGGTACTCTCTGCCCCCCCCCCCCAAAACCAGAATAC 7705

Qy   3661 GACCTGGAGGTTGATAACATCATGTCTCCTCCAAATGTGTGCCCACGGCACGGATGCATCGGC 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7706 GACTTGGAGGTTGGATAACATCATGGGCTCCTCCCAAATGTCCTGGCTCGGCGCACGGAATCTGGC 7765

Qy   3721 AAAAGGGGTATACTACCCTCACCCGTGACCC 3749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7766 AAAAGGGGTGTACTACTATTCACCCGGTGACCC 7794
```

```
Qy  1501 GGAACATTCCCCACCAACGCATACACCACCAGGGCGCCCTGCGACGCGCAAAC 1560
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5547 GGAACATTCCCCATTAAACGGGTACACCACGGGCGCCCTTCCCGGCGGGCCCAAAT 5606

Qy  1561 TATTCCAGGGCGCGCTGGTCGGCGGCGGGGGGTTGGCGGCGGGTGGGGG 1620
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5607 TATTCTAGGGCGCTGTGTGGGCGGGGTGGCGTGTGAGGAGTACGGCGGGGTGGGG 5666

Qy  1621 GATTTCCACTACGTGACGAGCACATGACCACTGACACAGTAAAATGCCCGTGCCAGGTTCCA 1680
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5667 GATTTCCACTACGGTGACGGGCCCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5726

Qy  1681 GCCCCCGGAATTCTTCACAGAGATGGATGGGGTGCGGCTTCCGGCGGGTGTGC 1740
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5727 GCCCCCGAATTCTTCACAGAGAGTGGATGGGGTGCGGCTTGCACAGGTACGGCTTCCAGCGTGC 5786

Qy  1741 AAACCTTCCTTCCTTACGGGGAGGAGGGGTCACCATTCCAGTCGGCGGCCTCAACCAA TACCTCGGTTGGG 1800
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5787 AAACCCTTCCTTACGGGGGAGAGGGTCACATTCCTGTTCGGGGCTCAATCAATACCTCGGTTGGG 5846

Qy  1801 TCGGCAGTCTCCCATGGGGGAGGCCCGAACCCGGAGGATGTAGCAGTGGCTCACTTCCATGCTCACCGAC 1860
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5847 TCACAGCTTCCCATCCGAGCCCGGAACCCGGAGCCTAGCAGTGGCTCACTTCCATGCTCACCGAC 5906

Qy  1861 CCCTTCCCACATCACAGCAGAGAGACGGGCTAAGGCCAGGGGTCTCCCCCCTCC 1920
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5907 GGGCGGGGGACATTACCCGGTGGGAGTCAGAGAACAAGGGTAGTACGCTAAGCCGTGGGACTCTTTCGAC 5966

Qy  1921 TCGGCCAGGTCTTCAGCTAGCCAGTTGTCTGGCCTTCCTTCGGAAGGCGACATACATTACC 1980
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  5967 TTGGCCCGCTCATCAGCTTAGCCACTGTCTCTGCGCCTTCCTTGGAAGGCAACATGCCAGTACC 6026

Qy  1981 CAAAATGACTTCCCAGACCGCTGACCTCATCGAGACCCATCCTCTGTGTGCCGGCCATGAGATG 2040
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6027 CGTCATGACTCCCCCGGACGGCTGACCTCATCGAGCCCAACCTCTGTGGGCGGGGCCAGGAGATG 6086

Qy  2041 GGGCGGGGACATTACCCGGTGGGAGTCAGAGAACAAGGGTAGTAATCCTGGACTCTTTTCGAC 2100
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6087 GGGCGGGAACATCACCCGCGGTCAGAAAATAAGGTAGTAATTTGGACTCTTTCGAG 6146

Qy  2101 CCGGCTCCGAGCGGAGGGAGTGAGGGAGGAGTGTCGGGGACGATCCTGCGGAAA 2160
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6147 CCGCTCCAAGCGGAGGGATGAGAGGGAAGTATCCGTTCGGGGGGGAGATCCTGCGGGAGG 6206

Qy  2161 TCCAAGAAATTCCCACCAGCGGATGGCCCGCATGGGGCACGCCCGGATTACAAACCCTCCGCTG 2220
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6207 TCCAGGAAATTCCCTCGAGCGGATGCCCATATGGGGCACGCCCGGATTACAAACCCTCCACTG 6266

Qy  2221 CTGGGAGTGCCTGGGAAAGGCCCCGGGGACTACGTCCCCTCCCCTCCAGTGGTACATGGGGTTGCCCACTGCCA 2280
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6267 TTAGAGTCCTGGGAAGGACCCGGACTACGTCCCCAGTGGTACACGGGGTGTCCATTGCCG 6326

Qy  2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGGAGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6327 CCTGCCAAGGCCCCTCCCGATACCACCTCCACGGGAGGAAGAGGGCGTCACCTTGGTACCCGG 6386

Qy  2341 TCCACCGTGTCTTCTGCCCTGGCCGGAGCTTGCCACACAAGGCTTTCGGGTTCCGAACCCG 2400
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6387 TCTACCGTGTTCTTCTGCCCTTGGCGCGCACCTTCGGCGACAGGGAG 6446

Qy  2401 TCGGCCGTGACAGCGGCAACCGCCCCCTCCTGACCGACGAGGGGCGGGCGGA 2460
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6447 TCGGCCGTGACAGGCACGGCGCTTGGCGGGGCTCGCGACGACGAGGAGGAGGA 6506

Qy  2461 GCAGGATCTGACGTTGAGTCGGTCGGTATTCCTCGATGCCCGGGGGGGGGAC 2520
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6507 GCGGGATCCGAGGGTTGAGTCGTGAGTACTTCCTCCATGCCCAAGAGGGGGAT 6566

Qy  2521 CCCGAACTCTCAGCGGACGGGGGTCTTGGTGCTACCGTGAGTGAGGACGTCGTC 2580
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  6567 CCCGAATCTCAGCGGACGGGGGTCTTGGTTCTACCCGTAAGCGACGAGGGGGTTCGTC 6626

Qy  2581 TGCTGCTCGGATGTCCTACACATGGACAGCGCCGCTCTGATCACGCCATGCGCGGCTGCGGAGGAA 2640
            |||||||||||||||||||||||||||||||||||||||||||||||

```

```
Db  6627 TGCTGGCTCGATGTCCTACACATGGACAGCGCCCCCTGATCACGCGCATGGCTGCGGAGGAA 6686
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6627 TGCTGCTCGGATGTCCTGCATGTCCTCACACAGTGACAATGCTGCGGAGGAA 6686

Db  2641 AGCAACGTGCCCATCAAACGCGGTTGAGCAACTCTTTGCCTCGGTCCTCCACCAACATGGTCTAC 2700
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2641 ACCAAGCGTGCCCATCAAATGCACTGAGCAACTCTTTCACCACAACTTGGTCTAT 2700

Db  6687 ACCAAGCTGCCCCATCAATGCACTGAGCAACTCTTTGCTCCGGTCCTCCACCAACATGGGTCTAT 6746
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6687 GCTACCACACATCCCGCGACGCGGCAGCCAGCGCAGGGCGGAGGGCGGGCGGAGGAAGGTTCACCTTTGGGCAA 6746

Db  2701 GCTACCACGATCCCGCGCACGCGGCAGCCAGCGCGGCGGCGGCAGAAGAAGAAGGTTCACCTTTGACGAAGTGCAG 2760
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2701 GCTACCAACATCTCGGCAGGCGGCAAGCCTGGCGGCGAGCGGAGAGAAGAAGGTCACCTTTGACGACTGCAG 2760

Db  6747 GCTACCAACATCTCGGCAGGCGGCAAGCCTGGCGGCGGAGCGGAGAGAAGGTCACCTTTGACAGACTGCAG 6806
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6747 ATCCTGGACGATCACTACCAGGACACGGTGCTCAAGGAGGATGAAGGGGAGGTCGAAGGCGGTCCCACAGTT 6806

Db  2761 ATCCTGGACGATCACTACCAGGACGCGGTGCTCAAGGAAGACGAGGAGGGGAGGTCGAAGGCGGTCCCACAGTT 2820
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2761 GTCCTGGACGGACCACTACCGGGGCGGTGCTCAAGGAGGATGAAGGGGAAGGGCTCCCACAGTT 2820

Db  6807 GTCCTGGACGGACACCACTACCGGGCGGTGCTCAAGGAGGATGAAGGGGAAGGGCTCCCACAGTT 6866
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6807 ATCCGGCTCCGTGTGGGAGGACTTGTTGGGAGGACACTGAAACACCAATTGACACCACCATC 6866

Db  2821 AAGGGCTAAGCTTCTTATCATGTTAGTAGTAGCGGACGACGCGGGCCCCACCCCACATTCGGCGCCAAA 2880
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2821 AAGGGCTAAGCTTCTTATCATGTTAGTAGTAGCGGACGACGCGGGGCCCCACATTCGGCGCCAAA 2880

Db  6867 AAGGGCTAAACTTTCTATCCGTGGGAAGAAGGCCTTTGTAAGCTACGGCCCCACATTCGGCGCCAGA 6926
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6867 TCTAAAATTTGGCTATGGGGGCAAAGGACGCTCCGGAACCTTATCCAGCGAAGGGCCATTAACCAC 6926

Db  2881 TCTAAAATTTGGCTATGGGGGCAAAGGACGCTCCGGAACCTTATCCAGCGAAGGGCCATTAACCAC 2940
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2881 TCTAAATTTGGCTATGGGGGCAAAGACGTCCGGAACCTTATCCGCAAGGGCCGTTAACCAC 2940

Db  6927 TCTAAATTTGGCTATGGGGGCAAAGACGTCCGGAACCTTATCCGCAAGGGCCGTTAACCAC 6986
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6927 GTCCCTGGACGGACGACCACTACCGGGGCAGCACTGAAACTGACACCACCACTC 6986

Db  2941 ATCCGGCTCCGTGTGGGAGGACTTGTTGGGAGGACACTGAAACACCAATTGACACCATC 3000
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2941 ATCCGGCTCCGTGTGGGAGGACTTGTTGGGAGGACACTGAAACACCAATTGACACCACCATC 3000

Db  6987 ATCCGGCTCCGGTGTGGGAGGACTTGCTGGGAAGGACCCAATTGACACACCAATTGACAGAGTT 7046
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  6987 GTCCCTGGACGGACGACCACTACCGGGGCAGCACTGAAACTGACACCACCACTC 7046

Db  3001 ATGGCAAAAAATGAGGTTTTCTGCCTCCAACCAGAGGACCAGCATCGC 3060
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3001 ATGGCAAAAATGAGGTTTTTCTGCGTCCAACCAGAGGAGCAGCAGCTTCGC 3060

Db  7047 ATGGCAAAAAATGAGGTTTTCTGCGTCGGTCCAACCGAGACGAAGCAGCCAGTCGC 7106
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7047 CTTATCGTGTTCCCAGACTTGGGGTGGGTCCGGTGGGTGGAGAAAATGGCCCTTCTATGACGTG 7106

Db  3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGGTGGTGGGGAGAAAATGGCCCTCTATGACGTG 3120
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGGTGGTGGGGAGAAAATGGCCCTCTATGACGTG 3120

Db  7107 CTTATCGTATTCCCAGATTTCCCCGCTTGTTTGACTCAAGCTATCCCTTTACCCTCTATGTG 7166
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7107 CTTATCGTATTCCCAGATTTCCCCGCTTGTTTGACTCAAGCTATCCCTTTACCATGTG 7166

Db  3121 GTCTCCACCCTCCCCTCAGGCGCTGTGAGGGGCCTGGATTCGCACTCTTTTCGAC 3180
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3121 ATTTATCAATGTTGTCACTTGGCCTCCACTTCGGCCTGCAAAGCTCCAGCTGCAAAGCTGCTCACAGAG 3180

Db  7167 GTCTCCACCCTCCCCTCAGGGCGTGATGGGGGCTCTTCATTACGGGATTCCAAATACTCTCCTGGA 7226
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7167 AATCTACCAAATGTTGTAACTTGGCCGGAACTTCGGCCTGCAAAGCTCCAGCTGCAAAGCTCCAGCTGCTCAGGAC 7226

Db  3181 CAGCGCGGTCGAGTTCCTGGTGAAACCCTGAAAATCAAAAAGGGCGGCCGAATCCTGCTCCGGG 3240
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3181 CGGGCGGGTCGAGTTCCTGGTTGAAACCCTGAAAATCAAAAAGGGCAGCGAGACTTTGCG 3240

Db  7227 CAGCGGTCGAGTTCCTGGTGAAATGCCTGGAAAGCAGACCTTATGCCTTCGCA 7286
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7227 CGGGCGGGTCGAGTTCCTGGTGAAATGCCTGGAAAGCAGACCTTATGCCTTCGCA 7286

Db  3241 TATGACACCCGCGTGTTTGACTCAACAGTCACTGAGATGACATCCGTGTTCGTGGAGGGTCA 3300
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3241 TATGACACCCGCGTGTTTGACTCAACAGTCACTGAGAATGACATCCGTGTTCGTGGAGGGTCA 3300

Db  7287 TATGACACCCGCGTGTTTTTGACTCAAGGGTCAACATCCTCAAAGGTCATGGCCCTGTTGAGAGGTCA 7346
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7287 TATGACACCCGCTGTTTGGTTTTGACTCAACGGGTCAACACCCTCAACGGTCATTGGTGAGAGAGGTCA 7346

Db  3301 ATTTATCAATGTTGTCACTTGGCCTCCACTTCGGCCTGCAAAGCTCCAGCTGCCA 3360
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3301 ATTTATCAATGTTGTCACTTGGCCTCCACTTCGGCCTGCAAAGCTCCAGCTGCCA 3360

Db  7347 AATCTACCAAATGTTGTAACTTGGCCGGAACTTCGGCCTGCGAAGCTGCACACAGGTGCTCACAGAG 7406
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7347 GCCTCGCCGCGGTGTGAGCTGCCAAAGCTCCACGCGCTGCTCGTGTGCGGGAG 7406

Db  3361 CGGCGGGTCGAGTTCCTGGTGAAATCAATTCAAAAAAGGCAGCGAGACTTGGCCCG 3420
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3361 CGGCTTTTATCGGGCGCAGTTGAACTTGGGCTCATGTAATTCTAAAGGGGCAAGAACTGCGGCCCGG 3420

Db  7407 CGGCTTTACATCGGGGCGCCCCGTTGGACTAAATTTCTAAAGGGCAGAACTGCGGCCCGG 7466
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7407 CGGCTTTACATCGGGGCGCCCCGTTGGACTAAATTTCTAAAGGGCAGAACTGCGGCCCGG 7466

Db  3421 TGCCGCGCGGGGCAGGGGTGCTGAGCGAGGTGCGGGTAATACCCTCACTACCTCACATGTTACTTGAAG 3480
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3421 TGCCGCGCGGGGCAGGGGTGCTGAGCGAGGTGCGGGTAATACCCTCACTACCTCACATGTTACTTGAAG 3480

Db  7467 TGCCCGGCGGAGCGGGGTTGTACTGGAACCAGTGCGGGTTAATACCCTACAGTGTTACTTGAG 7526
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7467 TGCCCGGCGGAGCGGGGTTGTACTGGAACCAGTGCGGGTTAATACCCTACAGTGTTACTTGAG 7526

Db  3481 GCCTCTGCACGCTGTTGAGCTGCAAAGCTGCACGATGCTCGTGTCGTGCGGGAGAC 3540
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3481 GCCTCTGCACGCTGTTGAGCTGCAAAGCTGCACGATGCTCGTGTCGTGCGGGAGAC 3540

Db  7527 GCCGCTGCCGGCCCTGTCGAGCTGCGGAAGCTCCAGGACTGCACGATGCTCGGTGTGCGGGAGAC 7586
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7527 GCCGCTGCCGGCCCTGTCGAGCTGCGGAAGCTCCAGGACTGCACGATGCTCGGTGTGCGGGAGAC 7586

Db  3541 GGGCCTTGTCGTTATCTGTGTGAGAGGGCGGGGGGAACCCAGGAGGACGAGGAGAGAGAGTC 3600
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3541 GGGCCTTGTCGTTATCTGTGTGAGAGGGCGGGGGGAACCCAGGAGGACGAGGAGAGAGAGTC 3600

Db  7587 GACCTTGTCGGTTATCTGTGTGAAAGCGCCCAAGAGACCCAAAGCGACGGGGCC 7646
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7587 GACCTTGTCGGTTATCTGTGTGAAAGCGCCCAAGAGACCCAAAGCGACGGGGCC 7646

Db  3601 TTCACGGAGGCTATGACTAGGTACTCTGCCGCCCGGGGGACCGGGGACACCAGAGATAC 3660
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3601 TTCACGGAGGCTATGACTAGGTACTCTGCCGCCCGGGGGACCGGGGACACCAGAGATAC 3660

Db  7647 TTCACGGCAGGCTTAGCATATAGAGTACTCTGCCCCCGGGGACCCCCCAAACCAGAAATAC 7706
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  7647 TTCACGGCAGGCTTAGCATATAGAGTACTCTGCCCCCGGGGACCCCCCAAACCAGAAATAC 7706

Db  2581 TGCTGCTCGGATGTCCTACACAGTGACAATGCTGCGGAGGAA 2640
            |||||||||||||||||||||||||||||||||||||||||||||||
Qy  3661 GACCTGGAGTTGATAACATCATGCTCCTCCAATGTGTCGTCGGCCACGAGCATCTGGC 3720

Qy  3661 GACCTGGAGTTGATAACATCATGCTCCTCCAAATGTGTCGTCCGCCGACGATGCATCTGGC 3720
```

```
Db  7707 GACTTGGAGTTGATAAACATCATGCTCTCCTCCAATGTGTCAGTCGCGGGCCACCGATGCATCTGGC 7766
              ||||||||
Qy  3721 AAAAGGGTATACTACACCTCACCCGTGACCC 3749
         ||||||||||   ||||||||||||||||||
Db  7767 AAAAAGGGGTGTACTATCTCACCCGTGACCC 7795

RESULT 8
US-09-539-601-13
; Sequence 13, Application US/09396601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 13
; LENGTH: 8649
; TYPE: DNA
; ORGANISM: Hepatitis C virus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(341)
; OTHER INFORMATION: construct I389/NS2-3'/wt
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (342)..(1193)
; OTHER INFORMATION: hepatitis c virus core-neomycin phosphotransferase
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (1202)..(1812)
; OTHER INFORMATION: internal ribosome entry site of
; OTHER INFORMATION: encephalomyocarditis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1813)..(8418)
; OTHER INFORMATION: hepatitis C virus NS2 - 5B
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (8419)..(8649)
; PUBLICATION INFORMATION:
; AUTHORS: Lohmann, Volker
; AUTHORS: Krner, Frank
; AUTHORS: Koch, Jan-Oliver
; AUTHORS: Herian, Ulrike
; AUTHORS: Theilmann, Lorenz
; AUTHORS: Bartenschlager, Ralf
; TITLE: Replication of subgenomic hepatitis c virus RNAs in a
; TITLE: hepatoma cell line
; JOURNAL: Science
; VOLUME: 285
; PAGES: 110-113
; DATE: 1999-07-02
; US-09-539-601-13

Query Match    87.8%;  Score 3291.4;  DB 4;  Length 8649;
Best Local Similarity   92.4%;  Pred. No. 0;  Mismatches 286;  Indels 0;  Gaps 0;
Matches 3463;  Conservative   0;  Mismatches 286;  Indels  0;  Gaps   0;
```

```
Qy   121 CAGGCCCCACCCTCCATCCATGGGATCAAAATGTGAGAGTCATCAGCCGACCGACCTGGAAGTTGTCATCATACGGGTCTCACG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4177 CAGGCTTCCAGCCTCCATCGTGTGGGACCAAAATGTGGGAAGTTGGGAAGTTGTCATACGGGTCTCACG 4236

Qy   181 CTGCGCGGGGCCCAACACCCTTGCTGTGATATAGGCTGGGAGCCGCCTCCAAAAACGAAGA 240
         ||||||||||   |||||||||||||||||||||||||||||
Db  4297 ACACACCCCATAACCAAAATACATCATGGCATGCAATGTCGCGTGACCTCGTCACG 4356

Qy   241 ACACACCCCATAACCAAAATTCATCATGGCATGCATGTGCAGCCGGACCGACCTGGTCATACGGGTCTCACG 300
         ||||||||||||||||||||||||||||||||||||||||||
Db  4297 ACACACCCCATAACCAAAATACATCATGGCATGCAATGTCGCGTGACCTCGTCACG 4356

Qy   301 AGCAACCTGGGGTCTGGTCGGGGGGGGGTCCTTGCCAGCTCGGCCGTATTGGCCTTGACAACA 360
         ||||||||||||||||||
Db  4357 AGCAACCTGGGGTCTGCTGGTCTAGGCGCGGAGCTCCTTGCCAGCTCGGCCGTAAGCAACA 4416

Qy   361 GGCACGTGGTCATTGTGGGTTAGGATCATCTTCTTGTCCGGGCCGGCTCCGGCTGTGCCGAC 420
         |||||||||||||||||||
Db  4417 GGCACGCGTGGTCATTGTGGGCCAGGATCATCTTTGTCCGGAAAGCCGGCCATCATTCCCGAC 4476

Qy   421 AGGGAAGTTCCTCTACCAGGAGTTCGATGAGATGGAAGAGAGGGCGGGGTCGCCACCTCCCTTAC 480
         ||||||||||||||||||
Db  4477 AGGGAAGTCCTTTACCGGGAGTTCGATGAGATGGAAGAGAGGTCGGCCTCACACCTCCCTTAC 4536

Qy   481 ATCGAGCAGGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAAGGGGCCTCGGGGTTGCTGCAG 540
         ||||||||||||||||
Db  4537 ATCGAACACAGGGGAATGCAGCTTGCCGCCGGAACAGAACAGAGGCAAATCGGGGTTGCTGCAA 4596

Qy   541 ACAGCCACCAAGCAAGCAACGCGGAGGCCGCGTGCTCTCCGGTGCGGTGAGTCCAAGTGGCGAGCCCTT 600
         ||||||||||||||||
Db  4597 ACAGCCACCAAGCAAGCAGGGGTGCTCTGCTCCCGGGGTGGGTGGGAAATCCAAGTGGCCGACCCTC 4656

Qy   601 GAGAGACCTTCTGGGGGGCGGAAAACATGTGGGAACTTCATCAGCGGCGGGATACAGTTGACTTAGCCAGGC 660
         ||||||||||||||||
Db  4657 GAAGGCCTTCGTGCGGGCGGCAGCATAGTGGGAGAATTTCATCAGCGCGGTATAATATTTAGCAGGC 4716

Qy   661 TTGTCCACTCTGCCTGGGAATCCCGGGATTGCGCATCACACTGATGATGGGCGGTTCACAGCCTCGTC 720
         ||||||||||||||||
Db  4717 TTGTCCAACTCTGCCTCGGCAACCCGCGGAGGCATCACACTGATGGCAATTCAGAGCCTCTATC 4776

Qy   721 ACTAGCCCGCTCACCACACCCAAATCTACCCCTCCTGCTTAAACATCCCTGGGGGGGATGGGTAGCC 780
         ||||||||||||||||
Db  4777 ACCAGCCCGGTCTCACACCAACATACCCCTCCTTTAACATCCTGTTAACTGGGGGGGGCC 4836

Qy   781 GCCCAACTCGGCCTCCCCCCAGTGCTTGCCTTCAGCTTTCGTAGGGCGGGCATTGCTGGTGCG 840
         ||||||||||||||||
Db  4837 GCCCAAACTTGCCTCCTCCCAGGGCTGCTTGTCTTCTGGGCCCGCATCGCTGCTGAGCG 4896

Qy   841 GCTGGTTGGGCAGCCATAGGCCTTGGGGAAGGTGCTTGTGGGACAATCTTGGCCGGGCTATGGGACA 900
         ||||||||||||||||
Db  4897 GCTGTTTGGGCCGCAGGCCTTGGGGAAGGTGCTTGGGGAGGTGCTTGTGGCAGTTTATGGGAGCA 4956

Qy   901 GGAGTTGGCAGGCGCGGCTTCGGTGCTTTAAGGTCATCAGCGCGGAAATGCCCTTCCACCGAG 960
         ||||||||||||||||
Db  4957 GGGGTTGGCAGGCGCGGCGCGGCTTTAAGGTCATGAGCGGCGGAGATGCCCTCCACCGAG 5016

Qy   961 GACCTTGGTTAACTTAACTTACTCCCTGCTATCCTCTCTCCTGGGGTGGGCCCCCAGTCGGTCGTCGGGTG 1020
         ||||||||||||||||
Db  5017 GACCTGGTTAACCTAACTCCTGCTATCCTCTGCTATCCTGGGGTGGGCCCCCAGTCGGTCGTCGGGTG 5076

Qy  1021 TGCGCAGCGGATACTCGGCGCACCTGGGGTCCAGGGGGAGGGGAGGGGTGTGGCAGTGGGGAATGAAC 1080
         ||||||||||||||||
Db  5077 TGCGCAGCGGGATACTGCGGCTCGGCCGCCCAGGGGGAGGGGAGGGCTGCGCCATGGCGTGGATGAAC 5136

Qy  1081 CGGCTGATAGGGTTGGCTTCGGCCTGGGGTAACCATGTTTCCCCCACCGCACTATGTGCCTGAG 1140
         ||||||||||||||||
Db  5137 CGGCTGATAGGGTTGCGGTTGGCTTGCGGGGGGTTAACCACGTTTCTCCCCCACTGTGCCTGAG 5196

Qy  1141 AGCGGACGGCCCGCAGCAGCGTCAGTCCAGATCCTTCCGGACCTTACTATATCACCCAACTGTTG 1200
         ||||||||||||||||
Db  5197 AGCGGACGGCTGCAGCGCACGTGTCACTCAGATCCTTCTCTAGTCTTACCATCACTCAGCTCGTCTG 5256

Qy  1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCAGCGCCCTGCTGCCCGGCGCTGGTGGCTA 1260
         ||||||||||||||||
```

```
                 |||||||| ||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db          5257 AAGAGGCTTCACCAGTGGATCAACGAGGACTGCTCCACGCCATGCTCCGGCTCGTGGCTA 5316

Qy          1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
                 || ||||||||||| |||||||||||| || ||| |||| |||||||||||||||||||||
Db          5317 AGAGATGTTTGGGATTGGATATGCACGGTGTTGACTGATTTCAAGACCTGGCTCCAGTCC 5376

Qy          1321 AAGCTCCTGCCGCGATTACCGGGAGTCCCCTTTTTCTCATGCCAACGTGGGTACAAGGGG 1380
                 ||||||||||||||||| |||||||||||||| |||||||| ||||||||||||||||||
Db          5377 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 5436

Qy          1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440
                 ||||||||||| |||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db          5437 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 5496

Qy          1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
                 ||||| ||||||||||||||||||||||||| ||||||| |||||||||||||||| |||||||
Db          5497 CATGTGAAAAACGGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5556

Qy          1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560
                 |||||||||||||| ||||| ||||||||||||||||||||||||||||| ||||||||
Db          5557 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 5616

Qy          1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620
                 ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5617 TATTCTAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 5676

Qy          1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACGTAAAATGCCCGTGCCAGGTTCCA 1680
                 |||||||||||||||||| ||||||||||||||||||||||| |||||||| ||||||||
Db          5677 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 5736

Qy          1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTGC 1740
                 |||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db          5737 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGCTCCAGCGTGC 5796

Qy          1741 AAACCTCTCCTACGGGAGGAGGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTTGGG 1800
                 ||||| ||||||||||||||||||||||||| ||||||||||||| |||||||||||||||
Db          5797 AAACCCCTCCTACGGGAGGAGGTCACATTCCTGGTCGGGCTCAATCAATACCTGGTTGGG 5856

Qy          1801 TCGCAGCTCCCATGCGAGCCCGAACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 1860
                 || ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          5857 TCACAGCTCCCATGCGAGCCCGAACCGGACGTAGCAGTGCTCACTTCCATGCTCACCGAC 5916

Qy          1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920
                 ||||||||||| || || |||||||||||||| ||||||||||||||| ||||||||||||
Db          5917 CCCTCCCACATTACGGCGGAGACGGCTAAGCGTAGGCTGGCCAGGGGATCTCCCCCCTCC 5976

Qy          1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980
                 ||||||||||| ||||||||||||| ||||||||||||||| |||||| |||| || ||||
Db          5977 TTGGCCAGCTCATCAGCTAGCCAGCTGTCTGCGCCTTCCTTGAAGGCAACATGCACTACC 6036

Qy          1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040
                 |   |||||| ||| |||||||||||||||||||||||||||||||||||||| ||||||
Db          6037 CGTCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATG 6096

Qy          2041 GGCGGGGACATTACCCGCGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
                 |||||| |||| ||||||||||||||||| || |||||||||||  |||||||||||||
Db          6097 GGCGGGAACATCACCCGCGTGGAGTCAGAAAATAAGGTAGTAATTTTGGACTCTTTCGAG 6156

Qy          2101 CCGCTCCGAGCGGAGGAGGATGAGCGGGAAGTGTCCGTCCCGGCGGAGATCCTGCGGAAA 2160
                 ||||||| |||||||||||||||| ||||||| ||||| |||||||||||||||||||
Db          6157 CCGCTCCAAGCGGAGGAGGATGAGAGGGAAGTATCCGTTCCGGCGGAGATCCTGCGGAGG 6216

Qy          2161 TCCAAGAAATTCCCACCAGCGATGCCCGCATGGGCACGCCCGGATTACAACCCTCCGCTG 2220
                 |||| ||||||||| | ||||||||||  ||||||||||||||||||||||||||| |||
Db          6217 TCCAGGAAATTCCCTCGAGCGATGCCCATATGGGCACGCCCGGATTACAACCCTCCACTG 6276

Qy          2221 CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGTGCCCACTGCCA 2280
                  | |||||||||||||| ||||||||||||||||||||||||||| ||||| ||| ||||
Db          6277 TTAGAGTCCTGGAAGGACCCGGACTACGTCCCTCCAGTGGTACACGGGTGTCCATTGCCG 6336

Qy          2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340
                 ||| | ||| ||||||| |||||||||||||||||||||||||| ||||| ||| |||||
Db          6337 CCTGCCAAGGCCCCTCCGATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGTCAGAA 6396

Qy          2341 TCCACCGTGTCTTCTGCCCTGGCGGAGCTTGCCACAAAGGCTTTCGGTAGCTCCGAACCG 2400
                 || ||||||||||||||| |||||||||| ||||||||| | ||||| ||||||||| ||
Db          6397 TCTACCGTGTCTTCTGCCTTGGCGGAGCTCGCCACAAAGACCTTCGGCAGCTCCGAATCG 6456

Qy          2401 TCGGCCGTCGACAGCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACGACGGCGGA 2460
                 |||||||||||||||||||||||||| ||| |||||||||| ||||||||||||||||
Db          6457 TCGGCCGTCGACAGCGGCACGGCAACGGCCTCTCCTGACCAGCCCTCCGACGACGGCGAC 6516

Qy          2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAC 2520
                 || ||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          6517 GCGGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAT 6576

Qy          2521 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGTGAGGAGGCCGGTGAGGACGTCGTC 2580
                 |||||||||||||||||||||||||||||||| || ||||||||  ||||||||||||||||
Db          6577 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTAAGCGAGGAGGCTAGTGAGGACGTCGTC 6636

Qy          2581 TGCTGCTCGATGTCCTACACATGGACAGGCGCTCTGATCACGCCATGCGCTGCGGAGGAA 2640
                 |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          6637 TGCTGCTCGATGTCCTACACATGGACAGGCGCCCTGATCACGCCATGCGCTGCGGAGGAA 6696

Qy          2641 AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTAC 2700
                 | ||||||||||||||| ||  |||||||||||||||||| ||||||||||||| |||||||
Db          6697 ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCCGTCACCACAACTTGGTCTAT 6756

Qy          2701 GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAA 2760
                 ||||| ||||| ||||||||||||| ||||||||||||||||||||||||||||||||||
Db          6757 GCTACAACATCTCGCAGCGCAAGCCTGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAG 6816

Qy          2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 2820
                  |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db          6817 GTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 6876

Qy          2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAAA 2880
                 |||||||| |||||||| || ||||||||||| ||||||||||||||||||||||||| |
Db          6877 AAGGCTAAACTTCTATCCGTGGAGGAAGCCTGTAAGCTGACGCCCCCACATTCGGCCAGA 6936

Qy          2881 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 2940
                 ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          6937 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCGTTAACCAC 6996

Qy          2941 ATCCGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000
                 ||||||||||||||| ||||||||| ||||||||||||| ||||||||||||||||||||||
Db          6997 ATCCGCTCCGTGTGGAAGGACTTGCTGGAAGACACTGAGACACCAATTGACACCACCATC 7056

Qy          3001 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGAGGAGGCCGCAAGCCAGCTCGC 3060
                 |||||||||||||||||||||||||||||||||||||  || ||||||||||||||||||
Db          7057 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAGGGGGGCCGCAAGCCAGCTCGC 7116

Qy          3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGTGTGTGCGAGAAAATGGCCCTCTATGACGTG 3120
                 |||||||| ||||||||| |||||||| |||||||||||||||||||||||| || || |||
Db          7117 CTTATCGTATTCCCAGATTTGGGGGTTCGTGTGTGCGAGAAAATGGCCCTTTACGATGTG 7176

Qy          3121 GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTACGGATTCCAGTATTCTCCTGGA 3180
                 |||||||||||||||||||| ||||||||||| || |||||||||||| || |||||||||
Db          7177 GTCTCCACCCTCCCTCAGGCCGTGATGGGCTCTTCATACGGATTCCAATACTCTCCTGGA 7236

Qy          3181 CAGCGGGTCGAGTTCCTGGTGAACGCCTGGAAATCAAAGAAGACCCCTATGGGCTTTGCA 3240
                 ||||||||||||||||||||||| |||||||||| | |||||   |||||||||||| |||
Db          7237 CAGCGGGTCGAGTTCCTGGTGAATGCCTGGAAAGCGAAGAAATGCCCTATGGGCTTCGCA 7296

Qy          3241 TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300
                 |||||||||||||||||||||||||| ||||||||||||||||||||||| |||||||||
Db          7297 TATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 7356

Qy          3301 ATTTATCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 3360
                 || || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7357 ATCTACCAATGTTGTGACTTGGCCCCCGAAGCCAGACAGGCCATAAGGTCGCTCACAGAG 7416

Qy          3361 CGGCTTTATATCGGGGGTCCCCTGACTAATTCAAAAGGGCAGAACTGCGGCTATCGCCGG 3420
                 |||||||| ||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db          7417 CGGCTTTACATCGGGGGCCCCCTGACTAATTCTAAAGGGCAGAACTGCGGCTATCGCCGG 7476
```

```
Query Match       87.8%;    Score 3291.4;    DB 4;    Length 11076;
Best Local Similarity   92.4%;    Pred. No. 0;    Matches 3463;
Matches 3463;    Conservative   0;    Mismatches 286;    Indels 0;    Gaps 0;

US-09-539-601-1
RESULT 9
US-09-539-601-1
; Sequence 1, Application US/09396010C
; Patent No. 6630343
; GENERAL INFORMATION:
;   APPLICANT: Bartenschlager, Ralf FW
;   TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;   FILE REFERENCE: all sequences
;   CURRENT APPLICATION NUMBER: US/09/539,601C
;   CURRENT FILING DATE: 2001-08-30
;   EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;   EARLIER FILING DATE: 1999-04-03
;   NUMBER OF SEQ ID NOS: 51
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11076
;   TYPE: DNA
;   ORGANISM: Hepatitis C virus
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (1)..(341)
;   OTHER INFORMATION: construct I389/Core-3'/wt
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (342)..(1193)
;   OTHER INFORMATION: hepatitis C virus core - neomycin
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1202)..(1812)
;   OTHER INFORMATION: internal ribosome entry site from
;       encephalomyocarditis virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1813)..(10845)
;   OTHER INFORMATION: hepatitis C virus polyprotein from core to
;       phosphotransferase fusion protein
;   FEATURE:
;   NAME/KEY: RBS
;   LOCATION: (1202)..(1812)
;   OTHER INFORMATION: nonstructural protein NS5B; parental sequence
;       without cell culture-adaptive mutations
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: (10846)..(11076)
```

(DNA sequence alignment between Qy and Db strands, positions 1–7805, omitted for legibility)

```
Db       7564 CGGCTGATAGCGTTCGCTTCGCGGGTAACCACGTCTCCCCCACGCACTATGTGCCTGAG 7623

Qy       1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTCTCCGACCTTACTATCACCCAACTGTTG 1200

Db       7624 AGCGACGCTGCAGCACGTGTCACTCAGATCCTCTCTAGTCTTACCATCACTCAGCTGCTG 7683

Qy       1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCACGCCCTGCTCCGGCTCGTGGCTA 1260

Db       7684 AAGAGGCTTCACCAGTGGATCAACGAGGACTGCTCCACGCCATGCTCCGGCTCGTGGCTA 7743

Qy       1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320

Db       7744 AGAGATGTTTGGGATTGGATATGCACGGTGTTGACTGATTTCAAGACCTGGCTCCAGTCC 7803

Qy       1321 AAGCTCCTGCCGCGATTACCGGGAGTCCCCTTTTTCTCATGCCAACGTGGGTACAAGGGG 1380

Db       7804 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 7863

Qy       1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440

Db       7864 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 7923

Qy       1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500

Db       7924 CATGTGAAAAACGGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 7983

Qy       1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560

Db       7984 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 8043

Qy       1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620

Db       8044 TATTCTAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 8103

Qy       1621 GATTTCCACTACGTGACGAGCATGACCACTGACAACGTAAAATGCCCGTGCCAGGTTCCA 1680

Db       8104 GATTTCCACTACGTGACGGGCATGACCACTGACAACGTAAAGTGCCCGTGTCAGGTTCCG 8163

Qy       1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGCTGCACAGGTACGCTCCGGCGTGC 1740

Db       8164 GCCCCCGAATTCTTCACAGAAGTGGATGGGGTGCGGTTGCACAGGTACGCTCCAGCGTGC 8223

Qy       1741 AAACCTCTCCTACGGGAGGAGGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTTGGG 1800

Db       8224 AAACCCCTCCTACGGGAGGAGGTCACATTCCTGGTCGGGCTCAATCAATACCTGGTTGGG 8283

Qy       1801 TCGCAGCTCCCATGCGAGCCCGAACCGGATGTAGCAGTGCTCACTTCCATGCTCACCGAC 1860

Db       8284 TCACAGCTCCCATGCGAGCCCGAACCGGACGTAGCAGTGCTCACTTCCATGCTCACCGAC 8343

Qy       1861 CCCTCCCACATCACAGCAGAGACGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCCTCC 1920

Db       8344 CCCTCCCACATTACGGCGGAGACGGCTAAGCGTAGGCTGGCCAGGGGATCTCCCCCCTCC 8403

Qy       1921 TTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCCTCGAAGGCGACATACATTACC 1980

Db       8404 TTGGCCAGCTCATCAGCTAGCCAGCTGTCTGCGCCTTCCTTGAAGGCAACATGCACTACC 8463

Qy       1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCATGAGATG 2040

Db       8464 CGTCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATG 8523

Qy       2041 GGCGGGGACATTACCCGCGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100

Db       8524 GGCGGGAACATCACCCGCGTGGAGTCAGAAAATAAGGTAGTAATTTTGGACTCTTTCGAG 8583

Qy       2101 CCGCTCCGAGCGGAGGAGGATGAGCGGGAAGTGTCCGTCCCGGCGGAGATCCTGCGGAAA 2160

Db       8584 CCGCTCCAAGCGGAGGAGGATGAGAGGGAAGTATCCGTTCCGGCGGAGATCCTGCGGAGG 8643

Qy       2161 TCCAAGAAATTCCCACCAGCGATGCCCGCATGGGCACGCCCGGATTACAACCCTCCGCTG 2220

Db       8644 TCCAGGAAATTCCCTCGAGCGATGCCCATATGGGCACGCCCGGATTACAACCCTCCACTG 8703

Qy       2221 CTGGAGTCCTGGAAGGCCCCGGACTACGTCCCTCCAGTGGTACATGGGTGCCCACTGCCA 2280

Db       8704 TTAGAGTCCTGGAAGGACCCGGACTACGTCCCTCCAGTGGTACACGGGTGTCCATTGCCG 8763

Qy       2281 CCTACTAAGACCCCTCCTATACCACCTCCACGGAGGAAGAGGACAGTTGTTCTGACAGAA 2340

Db       8764 CCTGCCAAGGCCCCTCCGATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGTCAGAA 8823

Qy       2341 TCCACCGTGTCTTCTGCCCTGGCGGAGCTTGCCACAAAGGCTTTCGGTAGCTCCGAACCG 2400

Db       8824 TCTACCGTGTCTTCTGCCTTGGCGGAGCTCGCCACAAAGACCTTCGGCAGCTCCGAATCG 8883

Qy       2401 TCGGCCGTCGACAGCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACGACGGCGGA 2460

Db       8884 TCGGCCGTCGACAGCGGCACGGCAACGGCCTCTCCTGACCAGCCCTCCGACGACGGCGAC 8943

Qy       2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAC 2520

Db       8944 GCGGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTTGAGGGGGAGCCGGGGGAT 9003

Qy       2521 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTGAGTGAGGAGGCCGGTGAGGACGTCGTC 2580

Db       9004 CCCGATCTCAGCGACGGGTCTTGGTCTACCGTAAGCGAGGAGGCTAGTGAGGACGTCGTC 9063

Qy       2581 TGCTGCTCGATGTCCTACACATGGACAGGCGCTCTGATCACGCCATGCGCTGCGGAGGAA 2640

Db       9064 TGCTGCTCGATGTCCTACACATGGACAGGCGCCCTGATCACGCCATGCGCTGCGGAGGAA 9123

Qy       2641 AGCAAGCTGCCCATCAACGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTAC 2700

Db       9124 ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCCGTCACCACAACTTGGTCTAT 9183

Qy       2701 GCTACCACATCCCGCAGCGCAAGCCAGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAA 2760

Db       9184 GCTACAACATCTCGCAGCGCAAGCCTGCGGCAGAAGAAGGTCACCTTTGACAGACTGCAG 9243

Qy       2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 2820

Db       9244 GTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTT 9303

Qy       2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAAA 2880

Db       9304 AAGGCTAAACTTCTATCCGTGGAGGAAGCCTGTAAGCTGACGCCCCCACATTCGGCCAGA 9363

Qy       2881 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCATTAACCAC 2940

Db       9364 TCTAAATTTGGCTATGGGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCGTTAACCAC 9423

Qy       2941 ATCCGCTCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATTGACACCACCATC 3000

Db       9424 ATCCGCTCCGTGTGGAAGGACTTGCTGGAAGACACTGAGACACCAATTGACACCACCATC 9483

Qy       3001 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAGAGGAGGCCGCAAGCCAGCTCGC 3060

Db       9484 ATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAGGGGGGCCGCAAGCCAGCTCGC 9543

Qy       3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGTGTGTGCGAGAAAATGGCCCTCTATGACGTG 3120

Db       9544 CTTATCGTATTCCCAGATTTGGGGGTTCGTGTGTGCGAGAAAATGGCCCTTTACGATGTG 9603

Qy       3121 GTCTCCACCCTCCCTCAGGCTGTGATGGGCTCCTCGTACGGATTCCAGTATTCTCCTGGA 3180

Db       9604 GTCTCCACCCTCCCTCAGGCCGTGATGGGCTCTTCATACGGATTCCAATACTCTCCTGGA 9663

Qy       3181 CAGCGGGTCGAGTTCCTGGTGAACGCCTGGAAATCAAAGAAGACCCCTATGGGCTTTGCA 3240

Db       9664 CAGCGGGTCGAGTTCCTGGTGAATGCCTGGAAAGCGAAGAAATGCCCTATGGGCTTCGCA 9723

Qy       3241 TATGACACCCGCTGTTTTGACTCAACAGTCACTGAGAATGACATCCGTGTAGAGGAGTCA 3300

Db       9724 TATGACACCCGCTGTTTTGACTCAACGGTCACTGAGAATGACATCCGTGTTGAGGAGTCA 9783
```

RESULT 10
US-10-029-907-6

Sequence 6, Application US/10029907
Patent No. 6706874
GENERAL INFORMATION:
APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
TITLE OF INVENTION: HEPATITIS C VIRUS
FILE REFERENCE: 13/083
CURRENT APPLICATION NUMBER: US/10/029,907
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,857
PRIOR FILING DATE: 2000-12-22
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 8638
TYPE: DNA
ORGANISM: HCV
FEATURE:
NAME/KEY: CDS
LOCATION: (1802)...(8407)
US-10-029-907-6

Query Match 87.7%;   Score 3288.2;   DB 4;   Length 8638;
Best Local Similarity 92.3%;   Pred. No.0;
Matches 3461; Conservative 0; Mismatches 288; Indels 0; Gaps 0;

Qy 1261
Db 5306
Qy 1321
Db 5366
Qy 1381
Db 5426
Qy 1441
Db 5486
Qy 1501
Db 5546
Qy 1561
Db 5606
Qy 1621
Db 5666
Qy 1681
Db 5726
Qy 1741
Db 5786
Qy 1801
Db 5846
Qy 1861
Db 5906
Qy 1921
Db 5966
Qy 1981
Db 6026
Qy 2041
Db 6086
Qy 2101
Db 6146
Qy 2161
Db 6206
Qy 2221
Db 6266
Qy 2281
Db 6326
Qy 2341

RESULT 11
US-10-029-907-7
Sequence 7, Application US/10029907
Patent No. 6706874
GENERAL INFORMATION:
APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
TITLE OF INVENTION: HEPATITIS C VIRUS
FILE REFERENCE: 13/083
CURRENT APPLICATION NUMBER: US/10/029,907
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,857
PRIOR FILING DATE: 2000-12-22
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 7
LENGTH: 8638
TYPE: DNA
ORGANISM: HCV
FEATURE:
NAME/KEY: CDS
LOCATION: (1802)...(8407)

Query Match   87.7%; Score 3288.2; DB 4; Length 8638;
Best Local Similarity   92.3%; Pred. No. 0;
Matches 3461; Conservative 0; Mismatches 288; Indels 0; Gaps 0;

```
              |||||||||||||||||||| |||||||||||||| |||||||||||| |||||||||||
Db         7646 TTCACGGAGGCTATGACTAGATACTCTGCCCCCCCTGGGGACCCGCCCAAACCAGAATAC 7705

Qy         3661 GACCTGGAGTTGATAACATCATGCTCCTCCAATGTGTCGGTCGCGCACGATGCATCTGGC 3720
                ||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db         7706 GACTTGGAGTTGATAACATCATGCTCCTCCAATGTGTCAGTCGCGCACGATGCATCTGGC 7765

Qy         3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
                |||||||| ||||| ||||||||||||||
Db         7766 AAAAGGGTGTACTATCTCACCCGTGACCC 7794

RESULT 12
US-10-029-907-25
; Sequence 25, Application US/10029907
; Patent No. 6706874
; GENERAL INFORMATION:
;  APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
;  TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
;  TITLE OF INVENTION:  HEPATITIS C VIRUS
;  FILE REFERENCE: 13/083
;  CURRENT APPLICATION NUMBER: US/10/029,907
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,857
;  PRIOR FILING DATE: 2000-12-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 8638
;   TYPE: DNA
;   ORGANISM: HCV
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1802)...(8407)
US-10-029-907-25

  Query Match             87.7%;  Score 3288.2;  DB 4;  Length 8638;
  Best Local Similarity   92.3%;  Pred. No. 0;
  Matches 3461;  Conservative    0;  Mismatches  288;  Indels    0;  Gaps    0;

Qy            1 TGGGAGGGCGTCTTCACAGGCCTCACCCACGTGGATGCCCACTTCCTGTCCCAAACAAAG 60
                ||||||||||||||| ||||||||||||||| | || ||||| ||| |||||||| || |||
Db         4046 TGGGAGGGCGTCTTTACAGGCCTCACCCACATAGACGCCCATTTCTTGTCCCAGACTAAG 4105

Qy           61 CAGGCAGGAGACAACTTCCCCTACCTGGTGGCGTACCAGGCTACTGTGTGCGCTAGGGCC 120
                ||||||||||||||||||||||||||||| || ||||||||||| ||||||||| ||||| 
Db         4106 CAGGCAGGAGACAACTTCCCCTACCTGGTAGCATACCAGGCTACGGTGTGCGCCAGGGCT 4165

Qy          121 CAGGCCCCACCTCCATCATGGGATCAAATGTGGAAGTGTCTCATACGGCTAAAGCCTACT 180
                ||||| ||||||||||||| ||||| |||||||||||||||||||||||||||||||||| 
Db         4166 CAGGCTCCACCTCCATCGTGGGACCAAATGTGGAAGTGTCTCATACGGCTAAAGCCTACG 4225

Qy          181 CTGCGCGGGCCAACACCCTTGCTGTATAGGCTGGGAGCCGTCCAAAACGAGGTCACCCTC 240
                |||| ||||||||| ||| ||||||||||||||||||||||| |||||||||||| ||   |
Db         4226 CTGCACGGGCCAACGCCCCTGCTGTATAGGCTGGGAGCCGTTCAAAACGAGGTTACTACC 4285

Qy          241 ACACACCCCATAACCAAATTCATCATGGCATGCATGTCAGCCGACCTGGAGGTCGTCACG 300
                ||||||||||||||||||| |||||||||||||||||||| || ||||||||||||||||||
Db         4286 ACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTCACG 4345

Qy          301 AGCACCTGGGTGCTGGTGGGCGGGGTCCTTGCAGCTCTGGCTGCGTATTGCTTGACAACA 360
                ||||||||||||||||| |||||| ||||| ||||||||||||||||||||||| ||||||||
Db         4346 AGCACCTGGGTGCTGGTAGGCGGAGTCCTAGCAGCTCTGGCTGCGTATTGCCTGACAACA 4405

Qy          361 GGCAGCGTGGTCATTGTGGGTAGGATCATCTTGTCCGGGCGGCCGGCTATTGTTCCCGAC 420
                |||||||||||||||||||| |||||||||||||||||  ||||||| ||  ||||||||
Db         4406 GGCAGCGTGGTCATTGTGGGCAGGATCATCTTGTCCGGAAGGCCGGCCATCATTCCCGAC 4465

Qy          421 AGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCGCGTCGCACCTCCCTTAC 480
                ||||||||||| |||| |||||||||||||||||||||||| || || ||||||||||||
Db         4466 AGGGAAGTCCTTTACCGGGAGTTCGATGAGATGGAAGAGTGTGCCTCACACCTCCCTTAC 4525

Qy          481 ATCGAGCAGGGAATGCAGCTCGCCGAGCAGTTCAAGCAAAAAGCGCTCGGGTTGCTGCAG 540
                ||||| |||||||||||||||||||||| || ||||| || || ||  ||||||||||||| 
Db         4526 ATCGAACAGGGAATGCAGCTCGCCGAACAATTCAAACAGAAGGCAATCGGGTTGCTGCAA 4585

Qy          541 ACAGCCACCAAGCAAGCGGAGGCCGCTGCTCCCGTGGTGGAGTCCAAGTGGCGAGCCCTT 600
                |||||||||||||||||||||||| |||||||||||||||||| |||||||||||  |||| 
Db         4586 ACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCGTGGTGGAATCCAAGTGGCGGACCCTC 4645

Qy          601 GAGACCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTACTTAGCAGGC 660
                ||  ||||||||||||||| || |||||||| ||||||||||||||||||| || |||||||||
Db         4646 GAAGCCTTCTGGGCGAAGCATATGTGGAATTTCATCAGCGGGATACAATATTTAGCAGGC 4705

Qy          661 TTGTCCACTCTGCCTGGGAATCCCGCGATTGCATCACTGATGGCGTTCACAGCCTCTGTC 720
                ||||||||||||||||| || |||||||| |||||||||||||||| |||||||||||| ||
Db         4706 TTGTCCACTCTGCCTGGCAACCCCGCGATAGCATCACTGATGGCATTCACAGCCTCTATC 4765

Qy          721 ACTAGCCCGCTCACCACCCAATCTACCCTCCTGCTTAACATCCTGGGGGGATGGGTAGCC 780
                || |||||||||||||||||||  |||||||||| |||||||||||||||||||||| |||
Db         4766 ACCAGCCCGCTCACCACCCAACATACCCTCCTGTTTAACATCCTGGGGGGATGGGTGGCC 4825

Qy          781 GCCCAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTAGGCGCCGGCATTGCTGGTGCG 840
                |||||||| ||||| ||||| |||||||| |||||||||||||||||||| ||||| |||
Db         4826 GCCCAACTTGCTCCTCCCAGCGCTGCTTCCGCTTTCGTAGGCGCCGGCATCGCTGGAGCG 4885

Qy          841 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGACATCTTGGCGGGCTATGGAGCA 900
                |||||||||||||||||||||||||||||||||||||| || ||||| || |||||||||
Db         4886 GCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGATATTTTGGCAGGTTATGGAGCA 4945

Qy          901 GGAGTGGCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGGCGAAATGCCCTCCACCGAG 960
                || |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| 
Db         4946 GGGGTGGCAGGCGCGCTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACCGAG 5005

Qy          961 GACCTGGTTAACTTACTCCCTGCCATCCTCTCTCCTGGTGCCCTGGTCGTCGGGGTCGTG 1020
                |||||||||||| |||||||||| |||||||| ||||| ||||| |||||||||||||||
Db         5006 GACCTGGTTAACCTACTCCCTGCTATCCTCTCCCCTGGCGCCCTAGTCGTCGGGGTCGTG 5065

Qy         1021 TGCGCAGCGATACTGCGTCGGCACGTGGGTCCAGGGGAGGGGGCTGTGCAGTGGATGAAC 1080
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         5066 TGCGCAGCGATACTGCGTCGGCACGTGGGCCCAGGGGAGGGGGCTGTGCAGTGGATGAAC 5125

Qy         1081 CGGCTGATAGCGTTCGCCTCGCGGGGTAACCATGTTTCCCCCACGCACTATGTGCCAGAG 1140
                ||||||||||||||||| ||||||||||||||| || |||||||||||||||||||| |||
Db         5126 CGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCTGAG 5185

Qy         1141 AGCGACGCCGCAGCACGTGTCACTCAGATCCTCTCCGACCTTACTATCACCCAACTGTTG 1200
                |||||||| ||||||||||||||||||||||||||    ||||| ||||| || ||| ||
Db         5186 AGCGACGCTGCAGCACGTGTCACTCAGATCCTCTCTAGTCTTACCATCACTCAGCTGCTG 5245

Qy         1201 AAGAGGCTCCACCAGTGGATTAACGAGGACTGCTCCACGCCCTGCTCCGGCTCGTGGCTA 1260
                |||||||| ||||||||||| ||||||||||||||||||||| ||||||||||||||||||
Db         5246 AAGAGGCTTCACCAGTGGATCAACGAGGACTGCTCCACGCCATGCTCCGGCTCGTGGCTA 5305

Qy         1261 AGGGATGTTTGGGACTGGATATGCACAGTTTTGGCTGACTTCAAGACCTGGCTCCAGTCC 1320
                || ||||||||||| ||||||||||| || ||| |||| |||||| ||||||||||||||
Db         5306 AGAGATGTTTGGGATTGGATATGCACGGTGTTGACTGATTTCAAGGCCTGGCTCCAGTCC 5365

Qy         1321 AAGCTCCTGCCGCGATTACCGGGAGTCCCCTTTTTCTCATGCCAACGTGGGTACAAGGGG 1380
                ||||||||||||||||| |||||||||||||| ||||||||| ||||||||||||||||| 
Db         5366 AAGCTCCTGCCGCGATTGCCGGGAGTCCCCTTCTTCTCATGTCAACGTGGGTACAAGGGA 5425

Qy         1381 GTCTGGCGGGGAGACGGCATCATGCAGACCACCTGCTCATGTGGAGCACAGATCACCGGA 1440
                ||||||||||| |||||||||||||| ||||||||| |||||||||||||||||||||||
Db         5426 GTCTGGCGGGGCGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACCGGA 5485

Qy         1441 CATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAGACCTGTAGTAACATGTGGCAT 1500
                ||||| ||||||| ||||||||||||||||| ||||||| |||||||||||||| |||||||
Db         5486 CATGTGAAAAACTGTTCCATGAGGATCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5545

Qy         1501 GGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCAAAC 1560
                |||||||||||||| ||||| ||||||||||||||||||||||||||||||| |||||||| 
Db         5546 GGAACATTCCCCATTAACGCGTACACCACGGGCCCCTGCACGCCCTCCCCGGCGCCAAAT 5605

Qy         1561 TATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTGGGG 1620
                ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
```

RESULT 13
US-10-029-907-4
Sequence 4, Application US/10029907
Patent No. 6706874
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; TITLE OF INVENTION: HEPATITIS C VIRUS
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 4
; LENGTH: 8643
; TYPE: DNA
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1802)...(8407)
US-10-029-907-4

Query Match   87.7%;   Score 3288.2;   DB 4;   Length 8643;
Best Local Similarity   92.3%;   Pred. No. 0;   Mismatches 288;   Indels 0;   Gaps 0;
Matches 3461;   Conservative 0;   Mismatches 288;   Indels 0;   Gaps 0;

[The remainder of the page consists of a dense nucleotide sequence alignment between Query (Qy) and Database (Db) sequences, with position numbers ranging from 1 to approximately 5800, shown in two columns.]

```
Db   5786 AAACCCCTCCTACGGGAGGAGGAGGTCACATTCCTGGTCGGGGCTCAAATCAATACCTGGTTGGG
          ||||  |||||||||||||||||||||||||||||||  ||||||||||||||||  |||| ||||
Qy        :

Db   5846 TCACAGCTCCCATGCTGGAGCCGAGGTAGCAGTGCTCACTTCACCGAC                5905
          ||  | |||||||||||| |||||||  ||| ||| |||||||||||||
Qy   1801 TCCGCAGCTCCCATGCTGAGCCGAGGTAGCAGTGCTCACTTCACGAC                 1860

Db   5906 CCCTTCCCACATTACGGCGGGAGAGACGGCTAGGGGCCAGGGGATCTCCCCCTCC         5965
          ||||||||||||  ||| ||| ||||| ||||||||||||| |||||||||||||
Qy   1861 CCCTTCCCACATCACAGACAGAGGCTAGGGGCCAGGGGTCTCCCCCTCC              1920

Db   5966 TTGGCCAGCTCTTCAGCTACCAGTTGTCTCGCGGCCTTCCTCGAAGGCGACATACATTACC  6025
          |||||||||  ||| || |||||| ||||||  ||||||||||||||||||||||  |||||
Qy   1921 TTGGCCAGCTCTTCAGCTACCAGTTGTCTCGCGGCCTTCCTCGAAGGCGACATACATTACC  1980

Db   6026 CGTCATGACTCCCCGGACCGCTGACCTCATCGAGGCGGCCGCAGCAGGTG            6085
          || || |||||||||||||| ||  |||||| | ||||||||||||| |||
Qy   1981 CAAAATGACTTCCCACGGACCGCTGACCTCATCGAGGCGGCCAACCTACC            2040

Db   6086 GGGGGGGAGCATTACCCGGCCTGGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC  6145
          |||||| |||||||||||| ||||||||| ||||||| ||| |||| ||||||| ||||||||
Qy   2041 GGGGGGGGACATCACCCGGCTGGGAGTCAGAAAAATAAGGTAGTAATTTTGGACTCTTTCGAG  2100

Db   6206 CTGGAGTCCTGGGAAGGCCCCGGACTAGTCCCTCCAGTGGTCACACTGCCA             6265
          |||||||| |||| |||| ||| |||||||||| |||| ||| ||||||||
Qy   2221 CTGGAGTCCTGGGAAGGCCCGGGAGTACAGGGGTGCCCACTGCCA                  2280

Db   6266 TTGGAGTCCTGGAAGGAACCCGGACTAGTCCTCCCAGTGTGTCACGGGGTGTCATTGCCG    6325
          ||||||||||| ||||||| || |||||||||||||| | ||| |||||| ||| ||||
Qy   2281 CCTACTAGCACCCCTCCTATACCACGGGAGGAGGAGGACAGTTGTTCTGACAGAA         2340

Db   6326 CCTGCCAAGGCCCCTCCGGATACCACCTCCACGGAGGAGGGAAGTATCCTGCGGAGG       6385
          |||||  |||| ||||||| ||||||| ||||||| |||| |||||||| |||||||
Qy   2341 TCCAACCGTGTCTTTCTGCCCTGCCCTGGGGGAGCTTGCCCACAGCTTTCGGGTRAGCTCCGAACCG  2400

Db   6386 TCTTACCGTGTCTTCTGCCTTGGCGAGCTCGCCCACAACAGACCTTGGCAGCTCCGAATCG   6445
          ||| |||||| || |||| ||||||||||| ||||| || ||||| |||||||| |||||
Qy   2401 TCGGCCCGTGCGACAGCGGCACGACGGCAACCGGCCCTCCTGACCAACCCTCCGACGACGGGGGA  2460

Db   6446 TCGGCCCGTCGACAGCGGCACGCGGCAACGGCCCTCTCCTCCGACAGCGACGGGGGAC      6505
          ||||||||| ||||||||||| |||||| ||||||| |||||||||||||||||||||
Qy   2461 GCAGGATCTGACGTTGAGTCGTATTCCTCCATGCCCCCTTTGAGGGGGAGCCGCGGGGGAC  2520

Db   6506 GGGGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCTTGAGGGGGTGAGCCGCGGGGAT  6565
          |||||||||||||||||||||| |||||||||||||||| |||||||| |||||||||||
Qy   2521 CCCGGAGTTCAGCGGACGGTCTTGGTCTCACGTGAGTGAGGCGGGGTTGAGGACGTCGTC   2580

Db   6566 CCCGGAGTCAGCGACGGGTCTTGGTCTCACGTGAGTGAGGCTAGTGGGACGTCGTC       6625
          ||||||||||||| || ||||||||||||||||||| ||||||||||||||||||
Qy   2581 TGCTGCTCGTCGATGTGCTACACATGACAGGCCATGGGCCTGCGGAGGAA            2640

Db   6626 TGCTGCTTCGGATGTCCTACACATGGGCCCATGCCGCTGCGGAGGAA                6685
          |||||||| || ||| |||||||| ||| || |||| ||||||||||
Qy   2641 AGCAAGCTGCCCATCAAGCGGTTGAGCAAACTCTTTGCTCGTCACCAGAACAATGGTCTAC  2700

Db   6686 ACCAAGCTGCCCATCAACAGCTCCTTTGCTCGTCACCAGAACAATGGTCTGTTCTAT      6745
          |||||||||||||| |  | || |||||||||||||||||||||||||| || |||
Qy   2701 GCTACCACCATCCCGCGCGGCGCAGCCGGACGGGAAGAAGGTCCGGGAACCTATCCGACTGCAA  2760

Db   6746 GCTACAACATCTCGGCAGGGCAGCCCGGACTGGGAAGAAGGTCACGTTTGACACGACTGCAG  6805
          |||| ||||| ||| |||| |||||||||| |||||| |||| | ||||||||||||||
Qy   2761 ATCCTGGACGACTCACTACCAGGCAGGGCGTGCTGCTCGCACAGTTT               2820

Db   6806 GTCCTGGACGACCACTCGGGCGGGACCGTGCTGCTCAAGGAGAGATGAAGGCGGTGGGGAT  6865
          |||||||||||||||| |||||||||||||||||||||||||| |||
Qy   2821 AAGGCTAAGCTTCTATCAGTAGGAGGAAGCCTGCAAGCTCGACGCCCCCACATTCGGCCAAA  2880
```

```
Db   6866 AAGGCTAAACTTCTATCCGTCGGAGGAAGCCTTGTAAGCTCGACGCCCCCACATTCGGCCAGA  6925
          |||||||| |||||||| || ||||||||||| |||||||||||||||||||||||||||||

Qy   2881 TCTAAAATTTGGCTATCGGGGCCAAAGGCAGTTCCGGAAGCTACCTATCCGCAGAAGGCCAT  2940
          |||||||||| | |||||||||| ||| |||||||||| |||| ||||| || ||||||
Db   6926 TCTAAAATTTGGCTATCGGGGCCAAAGAGCGTCCGGAACCGTCCGGAACCTATCCGCAGAAGGCCAC  6985

Qy   2941 ATCCCGCTCCCGTGTGGGAGGACTTGTTGGAAGACACTGAAACACCAATC             3000
          ||||| ||||| |||||| |||||| |||| ||| ||||||||||||||
Db   6986 ATCCCGCTCCCGTGTGGGAGGACTTGCTGGAAGACACTGAAACACCAATC             7045

Qy   3001 ATGGCGAAAAAATGAGGTTTTTCTGCGTTCCAACCAGAGGAGGAGGACGCAAGCACTCGC  3060
          |||||| |||||| | ||||||| ||||||||||||||||||||||||||||| ||||||
Db   7046 ATGGGCAAAAAATGAGGTTTTTCTGCGTTCCAACCAGAGGAGGAGGACGCAAGCACTCGC  7105

Qy   3061 CTTATCGTGTTCCCAGACTTGGGGGTCCGGTGTCGTGTGGGTGGGAGAAATGGCCTTATGACGTG  3120
          |||| |||||||||||| |||||| |||||| |||||||||| ||||| |||||| ||| ||||
Db   7106 CTTATCGTTATTCCCAGATTTGGGGGTTCGGTGTTCGGTGTGCCGGAGAAAATGGGCCCTTTACGAGTG  7165

Qy   3121 GTCTCCACCCTCCCTCCAGGCTGTGTAGGGGGGTGAGTGGGGGTTCAGTATCTCCAGTGGA  3180
          |||||||| ||||||||||| ||||||||||| ||||| |||| |||||||| ||||||
Db   7166 GTCTCCACCCTCCCTCCAGGCCGCGTGATGGGGGTCTTCATACGGCTCCTCCTGGA       7225

Qy   3181 CAGCGGGTCGAGTTCCTGGTTGAAACGCCTGGAAATCAAAGAAGGCTTTGCA           3240
          || || || |||||||| ||||||| |||| ||| |||| ||| |||||||
Db   7226 CAGCGGGTTCGAGTTCCTGGTTGAAATCCCTGGAAATGCCTGGAAGAAAAGGGGCTTTGCA  7285

Qy   3301 ATTTATCATGTTGTTGGACTTTGGCGTCCCCCCGAAAGGCCACCATCTAGGGTTACGGTGGGGGT  3360
          |||||| |||| ||| ||| |||||| ||||| |||||||||| || ||| || || || |||
Db   7346 ATCTACCAAGTGTTGTCACTTTGGCCCCCGAAAGCCACCATCAAGGGTTACGGTGGTGGGGGT  7405

Qy   3361 CGGCTTTATATCGGGGGTCCCCGTTGACTTCAAAAGGGGCCAGAAGCTGCGGCCGG      3420
          |||| ||||| ||||||| ||||| |||| |||||||||| || |||||||| ||||
Db   7406 CGGCTTTACATCGGGGGTCCCCCTGACTAATTCTAAAAGGGGCCAGAACTGCGGCCCG  7465

Qy   3421 TGCCGCGAGCGGGGCGGTGCTGACTAGCTGGCGTACTAATACCCTCACATGTTACTTGAAG  3480
          |||||||||||||||||| ||||| ||||| |||| |||||||| || ||||| |||||||
Db   7466 TGCCGCGAGCGGGGCGGTGTACTGACGACCAGTGTAATATCCCTCACATGTTAACTTGAAG  7525

Qy   3541 GGCCTTGTCGTTATCGTGAGAGGCGCGGAGAACCCAGGAGGACGCGGGCGGCCTAGGAGTC  3600
          |||||||| |||| | |||| |||| ||||| ||||||||||||| |||||||||||||||
Db   7586 GACCTTGTCGTTATCGTGTGAAAGCGGCGGGAGAACCCAGGAGGACGCGGAGGCCTAGGAGTC  7645

Qy   3601 TTCACGGAGGCTATGACTAGGTACTCTGCCCCCCGCGGGGGACCACCAAAGCAGAATAC    3660
          |||||||||||||| |||| ||||| |||||||| ||||||||| |||||  |||||
Db   7646 TTCACGGAGGCTATGACTAGATACTCTGCCCCCCAAACCAAGCAGAATAC             7705

Qy   3661 GACCTGGAGTTGATAAACATCAGTGCTCCTCCAAATGTTGTCGGGTTCGCGGAGTCATCTGGC  3720
          |||||||||| |||||||||||||||||||||||||||| |||||||||||||||| ||||
Db   7706 GACTTGGAGTTGATAAACATCAGTGCTCCTCCAAATGTTGTCGGGTTCGCGGAGTCATCTGGC  7765

Qy   3721 AAAAGGGTATACACCTCACCCGTGACCC                          3749
          ||||||| || || |||||||||||||
Db   7766 AAAAGGGTGTACTATCTCACCCGTGACCC                         7794
```

```
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 8648
; TYPE: DNA
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1802)...(8407)
US-10-029-907-5

Query Match    87.6%;   Score 3285;  DB 4;  Length 8648;
Best Local Similarity 92.3%;  Pred. No. 0;  Mismatches 290;  Indels 0;  Gaps 0;
Matches 3459;  Conservative 0;  Mismatches 290;  Indels 0;  Gaps 0;
```

```
Qy      1 TGGGAGGGGGGTCTTCACAGCCCTCACCCACCACGTGGATGCCCACTTCCTGTCCCAACAAG 60
          ||||| |||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   4046 TGGGAGGAGACCTCTTTTACAGGCCCTCACCCACATAGACGCCCATTTCTTGTCCCAGACTAAG 4105

Qy     61 CAGGCAGGAGAGACAACTTCCCCTACCTGGTGGGCGTACCAGCTACTGTGTGTGCGCTAGGGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4106 CAGGCAGGAGAGACAACTTCCCCTACCTGGTAGCAATACCAGCGTACTGTGTGCGCCAGGGCT 4165

Qy    121 CAGGCCCCACCTCCATCATGGGATCAAAATGTGGAAGTGTCTCATACGGCTAAAGCCTACT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4166 CAGGCTCCACCTCCATCGTGGGACCAAAATGTGGAAGTGTCTCATAAGGCCTACG 4225

Qy    181 CTGCGCGGGGCCCGCCAAACCCCTTGCTGTGTATAGGCGCGGGGGGTTCTCAGCTCTCG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4226 CTGCACGGGCCCAACGCCCCTGCTGTATAGGCTGGGAGGCCGTTCAAAACGAGAGTTTACTACC 4285

Qy    241 ACACACCCCATAACCAAAATTCATCATGGGCATGCATTGCAGCCGACCTGGAGGTCGTCACG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4286 ACACACCCCATAACCAAAATACATCATGGGCATGCATGTCAGCCGACCTGGAGGTCGTCACG 4345

Qy    301 AGCACCTGGGGTGCTGGTGGGGGTCCTTGCAGCTCTCGGCTGTATTGCTTGACAACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4346 AGCACCTGGGGTGCTGGTAGGCGGGACTCCTTAGCACGTATTGCCTGACACACA 4405

Qy    361 GGCAGCGTGGTGTCATTGTGGGGTAGGATCATCTTTGTTGTCCGGAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4406 GGCAGCGTGGTGTCATTGTGGGCCAGGATCATCTTGTCCGGGAAAGCCGGCCATCATTCCCGAC 4465

Qy    421 AGGGGAAGTCCTCTACCAGGAGGTTCGATGAGAGTGGCGGCGCTCGGCACCTCACCTCCCTTAC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4466 AGGGGAAGTCCTTTACGGGGAGTTCATCAGGAGGATGGAGAGAGTGCGCGCCCCTCACCTCCCTTAC 4525

Qy    481 ATCGGAGGCAGGGAATGCAGCTCGCCGAGCCAGTTCAAGCAAAAGCCTCGGGGTTGCTGCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4526 ATCGGAACAGGGGAATCCAGCTCGCCCGAACACAGATTCAAAACAGAGCCAATCGGGTTGCTGCAA 4585

Qy    541 ACAGCCACCAAGCAAGCGAGGCCCCTGCTCCCCGTGGGTGGAGTCCAAGTCGGGGAGCCCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4586 ACAGCCACCAAGCAAGCGAGGCCCGGATCCAAGTCCCAAGTGGCGGACCCTC 4645

Qy    601 GAGACCCTTCTGGGGCGAAACACATGGGGAACTTCATCATCGGGGGATACAGTAGTTAGCCAGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4646 GAAGCCCTTCTGGGGGAAAGCCTCACATATGTGGAAATTTCATCAGCGGGGGATACAGCAGG 4705

Qy    661 TTGTCCACTCTGCCTGGGAATCCCGGCGAGTTGCATCACTGATGATGGGCCATTCACAGCCCTCTAIC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4706 TTGTCCACTCTGCCTGGGAATCCCGGCGAGTTGCATCACTGATGATGGGCCATTCACAGCCCTCTAIC 4765

Qy    721 ACTAGCCCGTCACCAATCTACCCCTCGCTTACTTAAACATCCTTGGGGGATGGGTAGCCGG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4766 ACCAGCCCGTCACCCAACATACCCTCTGTTTAACATCCTTGGGGGATGGGTGGCC 4825

Qy    781 GCCCAACTTCCTCCCCCCAGTGGCGCTGCTTCTCAGCTTTCGGTTTAGGCGGGGCGGGGGCTTGGAGGC 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4826 GCCCAACTTGCTCTCCCAGCGGCTGCTTCGCTTTCGGTAGGGGTAGGCGGGGCGATGGGCTGGGTGGAGGC 4885
```

```
Qy    841 GCTGTTGGGCAGCATAGGCCTTGGGAAGGTGCTTGTGTGGACAATCTTGGCGCGGGGGCTATGGAGCA 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4886 GCTGTTGGGCAGGACTAGGCCCTTGGGAAGGTGCTTGGGGATATTTTGGCAGGTTATGGAGCA 4945

Qy    901 GGAGTGGCAGGCGCGGCGCTCGTCTGGCCTTTTGCCTTTAAGGTCATTGAGCGGGCGGAAATGCCCTTCACCGAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4946 GGGGTGGCAGGGCCGCTCGTGTGGCCTTTTGGGAAGGGTCTTGTTGGGATATTTGGGCAGGTTATGGGAGCCA 5005

Qy    961 GACCTGGTTAACTTACTCCCCTTGCCATCCTCTCCTTCCTGGTGTCCCCTTGGTGGCTA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5006 GACCTGGGTTAACCTACTCCCCTGCTAATCCTCCTCCTGGTGTCCCCTTGGTGGCTA 5065

Qy   1021 TGCGCGAGCGATACTTGCGTCCGGCACCGTGGGGTCCAGGGGGAGGGGGGCTGTGCAGTGGAGGATGAAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5066 TGCGCGAGCGGATAACTGCGTCGGGCACCGTGGGGCCCAGGGGGCCTGTGCGAGTGGAGGGAGGAAC 5125

Qy   1081 CGGCTGAGTAGCGTTCGCCCTTCGCCGCGGGGGGTAAGCACGTTCCCCCACGCGCCACTGTTGCCTGAG 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5126 CGGCTGAGTAGGGTTCGGCCTTTGGGGGCAGCCCACCACCACTATTGTGTGCCTGAG 5185

Qy   1141 AGCGGACGCCCCGCAGCACGTGCTCACTGAGATCCTCTCCGGACCTTACTTATCACCCAAGCTGTTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5186 AGCGGACGCTGGCAGGCGTGTTCACTCAGATCCTCTCTAGTTGTTCCATCACTCAGGTGCTG 5245

Qy   1201 AAGAGGCCTCACCCAGTGGGATTAACGATAGGACTTGGCCTCGCCTGCTTGCGCTGGCTA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5246 AAGAGGCCTTCAACCAGTGGGACTCCCAGGGAGGGACTGCTCCGCCATGGCCTCCGGCTA 5305

Qy   1261 AGGGGATGTTTGGGGACTGGGATGGCACAGTTTTGGGGTGGGACTTCAAGACCTGGCTTCCAGTCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5306 AGAGGATGTTTGGGGGTAGGGGTATGGCACGGGTGTTGACTGATTTCAACAGACCTGGCCTCCAGT 5365

Qy   1321 AAGCTTCCCTGCCGCGGATTACCGGGACGGGGAGTCCCCTTTTCTCATGGCCCCAACGTGGGGTAACAAGGGG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5366 AAGCTTCGTGCCGCGGATTGCGGGACGGGGAGTCCCCTTTTCTCATGGCCCAACGTGGGGTACAAGGGA 5425

Qy   1381 GTCTCGCGCGGGGAGAGGGGCAGCCAGTCCCATGCGAGACCACCTGCTTCATGTGGGGACAGATCACCGGA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5426 GTCTCGCGCGGGGAGGGGCGACGGGCATCATGCAGAAACACCTGCCCATGTGGAGGCACACGATCACCGGA 5485

Qy   1441 CATGTCAAAAAACGGTTCCATGAGGATCGTTGGCCTAAGCACCTTAGTAAGCATGTGGCAT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5486 CATGTGAAAAAACTGTTCCAATGAGGATTCGTGGGGCCTAGGACCTGTAGTAACACGTGGCAT 5545

Qy   1501 GGAACAATTCCCCATCAAACGCATACACCACGGGACGGGGCCCCCTGCAACGGCCAAAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5546 GGAACAATTCCCCATTAAACGGGTACACCACCAGGACGGGGCCCCTGCAACGTGTGGGCCAAAAT 5605

Qy   1561 TATTCCAGGGCGCTGTCGGGCGGGGTCGGCTGTGAGGAGTACGTGGAGGTTACGGTGGGG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5606 TATTCCTTAGGGCGTCGTGGGCGGGGTTGGCTGTGAGGAGTTACGGTGGGGGG 5665

Qy   1621 GATTTCCACTACGTGACGAGCACATGACACTGACAACGTAAAATGCCGGTGCCAGGTTCCA 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5666 GATTTCCACTACGTGACGGGCATGACACTGACACGTCGGGCCGGTGCCGTGTCAGGTTCG 5725

Qy   1681 GCCCCCGAATTCTTCACAGAAGTGGATGGGGGTGCGCGGCTGACACAGTGACGGGCCGTGC 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5726 GCCCCCGGAATTCTTCACGAGAAGTGGATAGGGGGTGCGGCTGCCAGCGTACGGTCCAGCGTGC 5785

Qy   1741 AAACCTCTCCTACGGGGAGGAGGTCACATTCCCATTCCAACCAATACCCTGGTTGGG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5786 AAACCCCTTCCTACGGCGGAGGAGGTCACACTTCTCGGGGCTCAAATACCCTGGTTGGG 5845

Qy   1801 TCGCAGCTTCCCATGCCGGAGCCCGAACCGGATGTACAGTGTCGGGGTTCACTTCCACTTCACCGGAC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5846 TCACAGCTCCCCATGCCGGAGCCCGAACCGGAGTGCAGCTTCACTTCCAATCCTCACCGGAC 5905

Qy   1861 CCCTCCCACATCACAGCAGAGACGCGCTAACGCAGGGGGGTTCTCCCCCTCC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5906 CCCTCCCACAATTACGGCGGAGACGGGCTAAGGGGATCTCCCCCGGGGGTCTCCCCGGGG 5965
```

```
Qy   1921 TTGCCAGCTGCTTCAGCTAGCCACTTGTGTCTGAGGCGCCTTTCCTCGAGAGGCGACTACATTACC 1980
          |||||||||||||||||||||||||||||||| ||||||||||||||||||  ||||||
Db   5966 TTGGCAGCTGCTCATCAGCTTAGCCAGCTGTCTGCGCGCCTCCTTGAAGGCAACATGACACTACC 6025

Qy   1981 CAAAATGACTTCCCAGACGCTGACCTCATCGAGCGCCAACCTCCTTGTGTGGGCGGCCATGAGATG 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6026 CGTCATGACTTCCCCGGACGCTGACCTCATCGAGCGCCAACCTCCTTGGGGGGGCCATGAGATG 6085

Qy   2041 GGGCGGGGACATTACCGCGGTGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCTTTCGAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6086 GGGGGGGACATCACCGCGGTGGAGTCAGAGAAAATAAGGTAGTAGACTCTTTCGAG 6145

Qy   2101 CCGCTCCCAGCGGAGGGGGGATGACGGGGAAAGTGTTCCGTCCCGGGGGAGGATCCTCCGGAAA 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6146 CCGGCTCCAAGCGGAGGGGGGATGAGGGGGGAAGTGTATCCGTTCCGGGGGAGG 6205

Qy   2161 TCCAAGAAATTCCCACCACGGATGCCCCGGATTACAACCCTTCCGCTG 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   6206 TCCAGGAAATTCCCTCGACGGATCCCCATATGGGCACGCCGGATTACAACCCCTCCACTG 6265

Qy   2221 CTGGAGTCCTGGAAGGCCCCGACTACGTCCCTCCAGTGGTACATGGGGTCCCACTGCCA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6266 TTAGAGTCCTGGAAGGACCCGGAGTTGTCCGTCCCGGGGGGAGGATCCTCCGGAAA 6325

Qy   2281 CCTACTAAGACCCCTCTATACACCTCAGCGAGAGAAGACAGTTTGTTCTGACAGAA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6326 CCTGCCAAGGCCCCTCCGACTACCACCTCCACGGGAGGAGGAGGTTGTCCTGTTCAGAA 6385

Qy   2341 TCCACCACGGTGTCTTTCTGCCCCTGGCCGTGAGCTTCCGTGAGTGAGCTCCACCG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6386 TCTACCGTGTCTTCGCCTTGGGGAGCGTCTTGGTCTACCGTAAAGAAGCTTTCGGTAGCTCCGAACCG 6445

Qy   2401 TCCGGCTCGTGGACAGGGCACGGCAACCACCCCCTCCTCCGGACGACGGGGGA 2460
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   6446 TCGGCCCGTCGTCGACAGGGCACGGCGCAACGGCCCTCCTCGACGGAGCCGGAC 6505

Qy   2461 GCAGGATCTGACGTTGATGTCGTATTCCTCCCATGGCCCCATGCCCTTGAGGGGGGGAC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6506 GCGGGATCCGACGTTGAGTCGTACTCCCTCCATGGGGGGGGAGGAGGGAGGGGGAT 6565

Qy   2521 CCCGGATCTCAGCGACGCGGGGTCTTGGTCTCAGCCGTGAGTGAGGACGGTCGTC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6566 CCCGGATCTCAGCGACGGGGTCTTGGTCTACCGTAAAGCAGCTTAGGCTGAGGACGTCGTC 6625

Qy   2581 TGCTGCTCGATGTCCTACACGACGGGGCTCTGATCACGCGCTGCCGAGGA 2640
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   6626 TGCTGCTCGAATGTCCTACACATGGACAGGGCCCTTGATCAGCCATGGGCCTGCGGAA 6685

Qy   2641 AGCAAGCTGCCCCATCAAACGGCGTTGAGCGAACTCTTTTGCGTGCGTCAGACAATTGGTCTAC 2700
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   6686 ACCAAGCTGCCCATCAAACTGCCACTGAGCAACTCTTTGCTCCACCAACTGTTCTAT 6745

Qy   2701 GCTACCACATCCCGCCAGGACGCCAAGCCCAGCGAGAGAAGGGTCACCTTTGACAGACTGCAA 2760
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   6746 GCTACAACATCTCGCAGGACCTACCAGGCGCCAAGCCTGGGCACGGGCAAGAGCTGCAG 6805

Qy   2761 ATCCTGGACGATCACTACCAGGACGTGCTCAAGGAGAGAAGGCCAAGGCCGTCCACAGTT 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6806 GTCCTGGACGGACGACACTACGGGACGTGCTCAAGGAGCGGGAAGGCGTCCACAGTT 6865

Qy   2821 AAGGCTAAGCTTCTATCAGTAGAGAGCTGCGAGCGCGCCCCCCACATTCGGCCCAA 2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6866 AAGGCTAAACTTCTATCCGTGGAGAAGCCTGTAAGCTGCAGCGCCCCACACTTCGGCCCAGA 6925

Qy   2881 TCTAAAATTTGGCTATGGGGCCAAAGGACGTCCGGAAGCCTTATCCAGCAAGGCCCATTAAACCAC 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6926 TCTAAAATTTGGCTATGGGGCCAAAGGACGTCCGGAACCTTATCCAGCAAGGCCCGTTAAACCAC 6985

Qy   2941 ATCCGCTCTCGTGTGGGAGGGACTTGTTGGAGAACAGCTGAAACACCAACCAATCC 3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6986 ATCCGCTCTCGTGTGGGAAGGGACTTTCCTGGAGAGACACTGAGACACCAATTGACACCAACCATC 7045

Qy   3001 ATGGCAAAAAATGAGGTTTTCTCGGGTCCAACCAGAGAGGAGGAGCCGCGAAGCCAGCTCGC 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7046 ATGGCAAAAAATGAGGTTTTCTGCGGTCCAACCAGAGAGGAAGGGGCCGCGAAGCCAGCTCGC 7105

Db   3061 CTTATCGTGTTCCCAGACTTGGGGGTGTCGTGTGGCGGCGGCTCTATGACGTG 3120
          |||||||||||||||||||||||||||||||||||||||| |||||||||||
Qy   7106 CTTTATCGTATTCCCAGATTTGGGGGTTCGTGTGGCGGGAAAATGGCCCCTTTACGATGTG 7165

Qy   3121 GTCTCCACCCTCCCTCCAGGCTGTGATGGGGGTGGTCCTCGTGTACCGGATTCCAGTATTCTCTGGA 3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7166 GTCTCCACCCCTCCCCAGGGCCGTGATGGGTGGGCTCTTCATACGGGATTTCCAAATACTCTTCGGA 7225

Qy   3181 CAGCGCGGTCGAGTTCCTGGTGAAACGGCCTGGAAATCAAAGGAGACCCCTATGGGCTTGCA 3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7226 CAGCGGGGTCGAGTTCCTGGTGAAATGCCTTGGAAATGCCTTGGGCTTATGGGCTTCGCA 7285

Qy   3241 TATGACACCCGCCGCTGTGTTTTGACTCAACAGTCACTGAGAATGCATCCGTTGCAGGTCA 3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7286 CGGCTTTACATCGGGGCACCCCTGTGTTTGACTCAACGGGCACTGAGAATCGCGGCTCA 7345

Qy   3301 ATTTATCAAATGTTGTGACTTGGCCCCCCGGAAGCCAGAGCAGCGCCCATAAGGTGTCACAGAG 3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7346 ATCTACCAAATGTGTGACTTGGCCCCCCGGAAGGCCAGAGCACGGCCCATAAGGGTCGCTCACAGAG 7405

Qy   3361 CGGCTTATATCGGGGGGGGTCCCCTGACTAATTCAAAAGGGGCAGAAACTGCGGGGCTATCGCCGG 3420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7406 CGGCTTTACATCGGGGCACCCCTGTGTTTTGACTCAACGGGCACTGAGAATTGCGGGCTATCGCCG 7465

Qy   3421 TGCCCGCGGAGCGGGGGGTGCTGTGACGACTAGTTGCGTGAGGTGGCGGGGGGC 3480
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   7466 TGCCGGCGGCGAGCGGGTGTATCTGACGCAGTCCAGTGCGGGTAATACCCTTCACTTCTGAAG 7525

Qy   3481 GCCCTGCGACGCTGTCGGAGGTGCAAAGCTTCCAGGACTGCACGGATGCTGGCGAGAC 3540
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   7526 GCCGCTGCGGCCCTGTCGAGCTCGCGGAAGCTCCAGGACTCGACGATCTCGTATGCGGAGAC 7585

Qy   3541 GGCCTTGTCGTTATCTGTGGAGCGCGGGGGGAACCCAGGAGGGGGGGGGGGCCTACGAGTC 3600
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   7586 GACCTTGTCGTTATCTGTGGAAAGCGGGGGCCAAGAGAGGACGGGAGCGGGGGGCC 7645

Qy   3601 TTCACGGGAGGCTATGACTTAGGTACTCTGCCCCCCCGGGGGGGGGGGGGGGAATAC 3660
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   7646 TTCACGGACGGGCTATGACTTAGATATCTCTGCCCCCCTGGGGAGACCAGAGATAC 7705

Qy   3661 GACCTGGGAGTTGATTAACATCATGCTCCCCCAAATGTGTCGGGGTCGGGTGCATCTGGC 3720
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   7706 GACTTGGGAGTTGATTAACACATCATCTCCCTCCAAATGTGTCAGTCGGGTCGGCCATCTGGC 7765

Qy   3721 AAAAGGGTATACTACCTCACCCGTGACCC 3749
          |||||||||||||||||||||||||||||
Db   7766 AAAAGGGTGTACTATCTCACCCGTGACCC 7794
```

```
;       OTHER INFORMATION: construct I389/NS3-3'/5.1
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: (342)...(1193)
;       OTHER INFORMATION: hepatitis C virus core - neomycin
;       OTHER INFORMATION: phosphotransferase fusion protein
;     FEATURE:
;       NAME/KEY: RBS
;       LOCATION: (1202)...(1812)
;       OTHER INFORMATION: internal ribosome entry site from
;       OTHER INFORMATION: encephalomyocarditis virus
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: (1813)...(7770)
;       OTHER INFORMATION: hepatitis C virus nonstructural proteins NS3 - 5B
;       OTHER INFORMATION: of cell culture-adapted clone no. 5.1
;     FEATURE:
;       NAME/KEY: 5'UTR
;       LOCATION: (7771)...(8001)
; US-09-539-601-22

Query Match     87.5%;  Score 3281.8;  DB 4;  Length 8001;
Best Local Similarity   92.2%;  Pred. No. 0;  Mismatches 292;  Indels 0;  Gaps 0;
Matches 3457;  Conservative    0;  Mismatches 292;
```

[Nucleotide sequence alignment. Each block shows a query row (Qy) and a database row (Db) with match bars (|) between them, with start and end position numbers flanking each line.]

```
Qy      1    TGGGAGGGGCCGTCTTCACAGGCCCTCACCCACGTGGATGCCCACTTCCTGTCCCAAACAAG   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3409  TGGGGAGGAGGCGTCTTTACAGGCCCTCACCCACATAGACGCCCCATTTCTTGTCCCAGACTAAG  3468
```

[Subsequent Qy/Db alignment blocks continue with the following coordinate ranges:]

```
Qy   61 – 120        Db 3469 – 3528
Qy  121 – 180        Db 3529 – 3588
Qy  181 – 240        Db 3589 – 3648
Qy  241 – 300        Db 3649 – 3708
Qy  301 – 360        Db 3709 – 3768
Qy  361 – 420        Db 3769 – 3828
Qy  421 – 480        Db 3829 – 3888
Qy  481 – 540        Db 3889 – 3948
Qy  541 – 600        Db 3949 – 4008
Qy  601 – 660        Db 4009 – 4068
Qy  661 – 720        Db 4069 – 4128
Qy  721 – 780        Db 4129 – 4188
Qy  781 – 840        Db 4189 – 4248
Qy  841 – 900        Db 4249 – 4308
Qy  901 – 960        Db 4309 – 4368
Qy  961 – 1020       Db 4369 – 4428
Qy 1021 – 1080       Db 4429 – 4488
Qy 1081 – 1140       Db 4489 – 4548
Qy 1141 – 1200       Db 4549 – 4608
Qy 1201 – 1260       Db 4609 – 4668
Qy 1261 – 1320       Db 4669 – 4728
Qy 1321 – 1380       Db 4729 – 4788
Qy 1381 – 1440       Db 4789 – 4848
Qy 1441 – 1500       Db 4849 – 4908
Qy 1501 – 1560       Db 4909 – 4968
Qy 1561 – 1620       Db 4969 – 5028
Qy 1621 – 1680       Db 5029 – 5088
Qy 1681 – 1740       Db 5089 – 5148
Qy 1741 – 1800       Db 5149 – 5208
Qy 1801 – 1860       Db
```

```
Db   5209 TCACAGCTTCCCATGCGACGCCGGAACCCGGAGTGTAGCAGTGCTCACTTCCATGCTCACCGAC 5268
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   5209 TCACAGCTTCCCATGCGACGCCGGAACCCGGAGTGTAGCAGTGCTCACTTCCATGCTCACCGAC 5268

Db   1861 CCCTTCCCACATTCACAGCAGAGACGGGCTAAGCGCAGGCTGGCCAGGGGGTCTCCCCCTCC 1920
Qy   1861 CCCTTCCCACATTAGGGCGGGAGACGGGCTAAGCGCAGGCTGGCCAGGGGGATCTCCTCCCCCC 1920

Db   5269 CCCTTCCCACATTAGGGCGGGAGACGGGCTAAGCGCAGGCTGGCCAGGGGGATCTCCTCCCCCC 5328
Qy   5269 ...

Db   1921 TTCGGCCAGCTCTTCAGCTTGTCTGCGCCTTCCTCGAAGGGCCAGATACATTACC 1980
Qy   1921 TTCGGCCAGCTCTTCAGCTTGTCTGCGCCTTCCTCGAAGGGCCAGATACATTACC 1980

Db   5329 TTGGCCAGCTCATCAGCTAGCCACCTGTCTGCGGCCTTCCTTGAAGGGCAACATGCACTACC 5388
Qy   5329 ...

Db   1981 CAAAATGACTTCCCAGACGCCTGACCTCATCGAGGGCCAACCTCCTGTGTGGCGGGCCATGAGATG 2040
Qy   1981 CGTTCATGACTCCCCGGACGGCTGACCTCATCGAGGGCCAACCTCCTGTGTGGGCGGGCAGGAGATG 2040

Db   5389 ...

Db   2041 GGGCGGGGACATTACCCGGGTGGGACTCAGAGAACAAAGGGTAGTAAT CCCTGGACTCTTTTCGAC 2100
Qy   2041 GGGCGGGGAACATCACCCGGGTGGGAGTTCAGAGAAAATAAGGTAGTAATTTTGGGACTCTTTCGAG 2100

Db   5449 ...

Db   2101 CCCGCTCCCAGCGCGAGGAGGATGAGCGGGGGAAGTGTCCGTCCCGGGGGGGAGATCCTTCCGGAAA 2160
Qy   2101 CAGCGCGGCGTCGAGTTCCCTGGGTGAAATCGGAAAGGGGACCCTAAGGGGGCTTTTGCA 2160

Db   5509 CCCCTTCCAAGCGGAGGAGATGAGAGGAAGTATCCGTTCCGGGGGGGAGGATTCCTGCGGAGG 5568

Db   2161 TCCAAGAAATTCCCACAGGGATACCCGGCCATGGGGCCACGCCCGGGATTACAAACCCTCCGCTG 2220
Qy   2161 TCCAGGAAATTCCCTCGAGCGGGATCCCATATGGGGCACGCCCGGGATTACAAACCCTCCACTG 2220

Db   5569 ...

Db   2221 CTGGAGTTCCTGGGAGGCCCCGGGACTACGTCCCCTCCCCTCCAGTGGTGACCACTGCCA 2280
Qy   2221 TTAGAGTCCTGGAGGACCCGGACTACGGTCCTTCCCAGTGGGTGTTCCATTGCCG 2280

Db   5629 ...

Db   2281 CCTACTAAGACCCCCTCCTATACCACCTCCACGGAGGAGGAAGCAGTTGTTCTGACAGAA 2340
Qy   2281 CCTGCCAAGGCCCTCCGATACCACCTTCACGGAGGAGGATCCTGGGGGAG 2340

Db   5689 ...

Db   2341 TCCACCGTGTCTTCTTGCCCTGGCCGAGCGTTGCCACGTGGTAGCTCCGAACCG 2400
Qy   2341 TGCCGCGCGGAGCGGGGTCGGTGTCGATGGGGTGAATACCCCTCCACAATGTTACTTGAAG 2400

Db   5749 TCTACCGGTGTCTTTCTGCCTTGGGGGAGCTCGCGCCACAGAGACCTTCGGCGCAGTCG 5808

Db   2401 TCCGGCCGTCGACAGCGGCGACGGAACACCCCCCTCCTGACCAACCCTCCTGACGCGACGGGGGCGA 2460
Qy   2401 ATCTACCAATGTTGTGACTTGGCCCCCCCCGGAAGCCACCGACGCCATAAGGTGCCTCACAGAG 2460

Db   5809 ...

Db   2461 GCAGGGATCTGAACGTTGAGTCGTAATTCCTCCAAGGGCCCCTTGAGGGGGGGAGCCCGGGGAC 2520
Qy   2461 CGGGCTTGTTGTTATCGGTGGGGTCCCCTGACTAATTCAAAAGGCGAACACTGCCCGGCTTATGCGGG 2520

Db   5869 GGGGGATCCGACGGTTGAGTCGTACTCCTCCCATGCCCCTTGAGGGGTGGAGCCCGGGGGAT 5928

Db   2521 CCCGATCTCAGCGACGGGGGTCTTGGTCTACGTGAGTGAGTCAGGCGGTCGGTCGTC 2580
Qy   2521 TGCCCGCGAGCGGGGTCGTGCTGTCTGACGACTAGCTGCGGTAATACCCTCCACCATGTTACCCGAGAAG 2580

Db   5929 CCCGGATCTTCAGCGACAGGGTTCTTGGTCTACCGTAAGCGAGGGGCTAGTGAAGGACGGTCGTC 5988

Db   2641 AGCAAGCTGCCCATCAACCGTTGAGCAACTCTTTGCTCGTCGGTCACCACAACATGGTCTAC 2700
Qy   2641 ACCAAGCTGCCCATCAATGCACTGAGCAACTCTTTGCTCGGTCACCACAAACTTGGTCTAT 2700

Db   6049 ...

Db   2701 GCTACCACATCCCGCAGCCAGCGCAGAGAGAGGT CACCTTTTGACAGACTGCAA 2760
Qy   2701 GCTACAACATCTCGCAGCCAAGCTGGCAGAGAAGAGGTCACCTTTGACAGAGCAC 2760

Db   6109 ...

Db   2761 ATCCTGGACGCGATCACTACCACAGGACGTGCTGCTCAACGGGAGGATGAAGGGCGAAGGGCTCCACAGTT 2820
Qy   2761 ...

Db   2821 AAGGCTAAGCTTCTATCAGTAGAGGAAGCCTGACGCCCCCCACATTCGGCCAAA 2880
Qy   2821 ...

Db   6229 AAGGCTAAACTTCTATCCGTGGGAGGAAGCCCTGTGAAGCTGAACGGCCCCCACATTCGGCCAGA 6288

Db   2881 TCTAAATTTGGCTATGGGGGCCAAAGGACGTCCGGAACCTATCCAGCAGGCCAAGGCCGTTAACCAC 2940
Qy   2881 ...
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: February 21, 2005, 10:29:03 ; Search time 11361 Seconds
                                              (without alignments)
                                              12564.124 Million cell updates/sec

Title: US-09-664-363-22
Perfect Score: 3750
Sequence: 1 TGGGAGGGGCGTCTTCACAGG..........ACTACCTCACCGTGACCCG 3750

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 34239544 seqs, 1903213470 residues
Total number of hits satisfying chosen parameters: 68479088

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :   EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_htc:*
               4: gb_est3:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_gss1:*
               9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	49.2	1.3	925	9	CNS0091P	AL053013 Drosophil
2	48	1.3	552	7	CNS02R3T	CF821532 Drosophil
3	48	1.3	856	7	CNS04VB9	CF824031 EST701413
4	47.6	1.3	925	9	CNS04IIT	CF824624 EST702006
5	47.6	1.3	978	7	CNS0091P	AL053013 Drosophil
6	47	1.3	635	9	CNS0091P	CF824954 EST702336
7	47	1.3	635	9	Leishmani	BX905657 Leishmani
8	47	1.3	855	7	LBAF018C06	BX545484 Leishmani
9	47	1.3	909	7	CO01S154	C001S154 EST785536
10	47	1.3	914	7	CF823313	CF823313 EST700695
11	47	1.3	923	7	CF826948	CF826948 EST704330
12	47	1.3	925	7	CF823378	CF823378 EST700760
13	46.2	1.2	925	7	CF825990	CF825990 EST703372
14	45.8	1.2	645	7	CNS01213	AL101589 Drosophil
15	45.6	1.2	881	7	CNS01213	AL101589 Drosophil
16	45.6	1.2	506	5	BP153627	BP153627 BP153627
17	45.4	1.2	652	7	CN010969	CN010969 WHE3878_D
18	45.4	1.2	753	9	CG855542	CG855542 ZMMBBc025
19	45.2	1.2	1101	9	CNS01720	AL107514 Drosophil
20	45.2	1.2	1087	9	BZ573831	BZ573831 mbh2_3381
21	44.4	1.2	1101	9	CNS017SY	AL108460 Drosophil
22	44.4	1.2	584	1	AV922850	AV922850 AV922850
23	44.4	1.2	1101	9	CNS017SY	CL945925 OsIFSB001
24	43.8	1.2	914	2	BF257968	BF257968 HVSMEf001
25	43.4	1.2	587	6	CA679773	CA679773 wlm4.pk00
26	43.2	1.2	966	9	CNS02R3T	AL210098 Tetraodon
27	43.2	1.2	971	9	CNS04VB9	AL308862 Tetraodon
28	43.2	1.2	1019	7	CNS04IIT	AL292286 Tetraodon
29	43	1.1	487	7	CV039737	CV039737 4137143 B
30	43	1.1	540	6	CD727100	CD727100 4031393 1
31	43	1.1	573	6	CD216293	CD216293 pgp2n.pk0
32	43	1.1	580	4	AV922834	AV922834 pgp1n.pk0
33	43	1.1	591	4	BI391789	BI391789 pgf2c.pk0
34	43	1.1	601	3	BM425602	BM425602 pgf2c.pk0
35	43	1.1	615	4	BM425623	BM425623 pgf2c.pk0
36	43	1.1	664	1	AJ454042B	AJ454042B pgmlc.pk0
37	43	1.1	665	1	AJ455557	AJ455557 pgmlc.pk0
38	43	1.1	667	1	AJ450427	AJ450427 AJ450427
39	43	1.1	673	7	CO507218	CO507218 GGEZEB201
40	43	1.1	674	1	AJ449106	AJ449106 AJ449106
41	43	1.1	690	1	AJ445947	AJ445947 AJ445947
42	43	1.1	695	4	BM486338	BM486338 pgmlc.pk0
43	43	1.1	698	5	BU481178	BU481178 603471787
44	43	1.1	704	6	CD763744	CD763744 GGEZLB100
45	43	1.1	727	6	CD764115	CD764115 GGEZLB100

ALIGNMENTS

RESULT 1
CNS0091P
LOCUS   CNS0091P   925 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION   AL053013.1 GI:4934461
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequence;
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES   Location/Qualifiers
     source   1..925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
ORIGIN
Query Match   1.3%; Score 49.2; DB 9; Length 925;
Best Local Similarity   13.9%; Pred. No. 0.34;

Matches 51; Conservative 161; Mismatches 154; Indels   0; Gaps   0;

```
Qy   761 TCCTGGGGGGGATGGGTAGCGCGCCAACTCGCTCCCCCAGTGTCGTTGCTTCAGCTTTCGTAG 820
Db   553 TTSSGGSGYGKGCSGSGGBSCSCCSCSCSCCBCCCSCSCCSSYCCSSSBSSSKCSS 612
Qy   821 GCGCGCGGCAATTGCTGGTGCGGCGTGTTGGGAGCATAGGCCTTGGGAAGTGCTTGTGGACA 880
Db   613 TSBSCSCCCSSKSVCGTSCSCSSSSSSSSSSTSSSSSTSKSSSSSGSSSSSSYTTS 672
Qy   881 TCTTGGCCGGGGCTATGGAGCAGGAGTGGGGCAGGCGCGCGCGCSSSSSSSSSSSSSSSSSSS 940
Db   673 KSTSASGSGSSWSAGGGSGTGSTSSSSGSTSTSSSSSSSSSSSSSSTSSSSSSSSS 732
Qy   941 GCGAAAATGCCCTCACCAGGAGGACCTGGGTTAAACTTACTCCCCSYSSSSSTSSSSSSSSSS 1000
Db   733 SSSTTSSSBBSCTSTSTSSSSSSSYSSSSSTSSSSSSSSSSSSSSSSSSSSSSSSV 792
Qy  1001 CCCTGGTGTCGTCGGGGGTGTGTCGGCCAGCGGATACTGCCCCSYSYSSSSYSSSSSSSSSSS 1060
Db   793 GTSSSSSDSTSTCSCSCCCYMCTCCSTYBMBCYTSTSCGSGSSSSGKKGGVTKCGCGCGSSS 852
Qy  1061 GGGCTGTGTGCAGTGGATTAGCCMCGGTTTGCCTCGCGGTGGGTAACCATGTTTCCC 1120
Db   853 TNGMBGTSSACSSSSSSSVSSSSKSSASSSSSGVSSNSSSASKSSSSSGSV 912
Qy  1121 CCACGC 1126
Db   913 SSGSGS 918
```

RESULT 2

```
LOCUS       CF824031/c
DEFINITION  EST701413    Coccidioides posadasii    552 bp    mRNA    linear    EST 01-APR-2004
            4 kb Coccidioides posadasii saprobic phase cDNA library, 2 to 4
            kb Coccidioides posadasii cDNA clone CIDAP05 5' end, mRNA
            sequence.
ACCESSION   CF824031
VERSION     CF824031.1    GI:4530088
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
  ORGANISM  Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Gardner,M.J. and Cole,G.T.
  TITLE     Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
  JOURNAL   Unpublished (2003)
COMMENT     Other ESTs: EST701412
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..552
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDAP05"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN
```

Query Match   1.3%; Score 48; DB 7; Length 552;
Best Local Similarity   50.4%; Pred. No. 0.63;
Matches 117; Conservative   0; Mismatches 115; Indels   0; Gaps   0;

```
Qy   797 CCAGTGCTGCTTCAGCTTTCGTAGCGCGCCGGCCAATGCTGGTGCGGCGTGTTGGCAGCATAG 856
Db   522 CCATGGCGCGGCCGTACTTGGGACACCGCGCCGCCGGCCTCGCGGCTATGAGACGAGAGCGCGC 463
Qy   857 GCCTTGGGAAGTGTGCTTGTGGACACATCTTTGGCGCGGCTATGAGACGAGAGGTGGCAGCGCGC 916
Db   462 CCATGGCGCCGCCGTGCTTGTGGCCATCATGTCCATGGCGCCGCCGGCCGTTAACTTAC 403
Qy   917 TCGTGGCCCTTTAAGGTCATGAGCGCGGCGGGAAATGCCCTCCACCAGGAGGACCTGGTTAACTTAC 976
Db   402 TTGTGGCGCGTCGTTGGTCCATGGCGCCGCCGGCCTCGGTGCGGTGCCCGTTAACTTAC 343
Qy   977 TCCCTCCACCATCATGTCCATGGCGCCGCCGGCCGTGCTTGTGCGGCGTGTCGTCGCAGC 1028
Db   342 TTGTGGCCCATCATGTCCATGGCGCCGCCGGCCGTGCTTGTGCGGCGTGTCGTCGCAGC 291
```

RESULT 3

```
LOCUS       CF824624/c
DEFINITION  EST702006    Coccidioides posadasii    856 bp    mRNA    linear    EST 01-APR-2004
            4 kb Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDAS39 5' end, mRNA
            sequence.
ACCESSION   CF824624
VERSION     CF824624.1    GI:45930681
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
  ORGANISM  Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 856)
  AUTHORS   Gardner,M.J. and Cole,G.T.
  TITLE     Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..856
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDAS39"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA
                     library, size fractionated cDNA 2 to 4 kb"
ORIGIN
```

Query Match   1.3%; Score 48; DB 7; Length 856;
Best Local Similarity   50.4%; Pred. No. 0.69;
Matches 117; Conservative   0; Mismatches 115; Indels   0; Gaps   0;

```
Qy   797 CCAGTGCTGCTTCAGCTTTCGTAGCGCGCCGGCCAATGCTGGTGCGGCGTGTTGGCAGCATAG 856
Db   791 CCATGGCGCGGCCGTACTTGTGACCGCGTCGTTGCCGGGCCGTGTTGGCGCAGCATAG 732
Qy   857 GCCTTGGGAAGTGCTTGTTGTGGACACATCTTTGGCGCGGCTATGGACGAGAGGTGGCAGCGCGC 916
Db   731 CCATGGCGCCGCCGTGCTTGTGGCCATCATGTCCATGGCGCCGCCGGCCGTGTTGGCGCGTCGTGC 672
```

RESULT 4

```
LOCUS       CNS0091P/c
DEFINITION  CNS0091P       925 bp    DNA    linear    GSS 03-JUN-1999
            Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

Query Match   1.3%; Score 47.6; DB 9; Length 925;
Best Local Similarity   15.3%; Pred. No. 0.9;
Matches   57; Conservative 164; Mismatches 148; Indels   3; Gaps   1;

```
Qy   1798 GGGTCGCAACTCCCATGCCGAGCCCGAACCGGATAGCAGTGCTCACTTCCATCTCACC 1857
          :||: :::|| ::|:  ::|||  ::|::||| :::::::  :||:|  || :||||
Db    914 SSBSCCSSSMSTSSSNSSBCCSSCCSSBBSSSSSSGGSSSSSGTSSACVKC 855

Qy   1858 GACCCCTCCCACATCACAGCAGAGACGGCTAGCGCCAGCGGGGTCTCCCCCC 1917
          :|||: :: ::::|  ::|||:|| |:|  :  ::|::||:|||:|:::::
Db    854 NASSSCCGCGCGMABCCMCSSSSSCCSASASARGVKVRASGGAGKRGGGSGASHSSSS 795

Qy   1918 TCCTTGGGCCAGCTCTTCAGCTAGCCAGTTG---TCTGCCCTTCCTCGAAGGCGGACATAC 1974
          | ||| ::|| :::||| |||:|:||::| |   | : ::|:|:::: :|::::||:
Db    794 ACBSSSSSCSASCWSASSSASSRSRSGGGAGSGGASSSRSSSSSSASAGSVVSSAS 735

Qy   1975 ATTACCCAAAATGACTTCCCAGAGAGCGCTGACCTCATCGAGGGGGCCAACCTCCTGTGCGGGCAT 2034
          :|: .| ::: :::| ::::|  : |||| ::||| ||||| ||||||: :||| ::||:|:::|
```

```
Db    734 SSSSSCCSSSVSCSSVVASSMCSSBBSSSSSASSSSSSSSSACASCSCCTWSCSCSTSA 675
Qy   2035 GAGATGGGGCGGGGGACATTACCCGCGTGGGAGTCAGAGAACAAGGTAGTAATCCTGGACTCT 2094
          ::|:::||| :::|:|:|:|  ||||||:||:|:|||||| |:||||:|:|| :|:|| |:|
Db    674 SMSAARSSSSSSSSSSSSSSCCSSSSSMSASSSSASSSSSSSSSSSGSACGBSMSSSGGSGSV 615

Qy   2095 TTCGACCCGGTCTCCGAGCCGAGGAGGGAGGAGGAGATGACGCCGGAAGTGTCCGTCCCGGGGAGGATCCTG 2154
          | |:||| ||:|||:|:|:||::|||| || || ||| :||: ::||||:|:: |:| :::|:||::|:: |:|
Db    614 SASSGMSSSVSSSGGRSSSGGGSGSGGGVGGSSGSSGSSSSSSVSCSSSGCMCRCSCSS 555

Qy   2155 CGGAAATCCAAG 2166
          |||  |:|:|:
Db    554 AAAAAASCVAAS 543
```

RESULT 5

```
LOCUS       CF824954      978 bp    mRNA    linear    EST 01-APR-2004
DEFINITION  CF824954 EST702336 Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDAU34 3' end, mRNA
            sequence.
ACCESSION   CF824954
VERSION     CF824954.1  GI:45931011
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
ORGANISM    Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 978)
AUTHORS     Gardner,M.J. and Cole,G.T.
TITLE       Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
JOURNAL     Unpublished (2003)
COMMENT     Contact: Gardner MJ
            Other ESTs: EST702337
            Email: gardner@tigr.org.
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org.
FEATURES             Location/Qualifiers
     source          1..978
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDAU34"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN
```

Query Match   1.3%; Score 47.6; DB 7; Length 978;
Best Local Similarity   50.0%; Pred. No. 0.91;
Matches 119; Conservative   0; Mismatches 119; Indels   0; Gaps   0;

```
Qy    783 CCAACTCGCTCCCCCCAGTGCTCGCTTCAGCTTTCGTAGCGCGCCGGCGTCGGCGGC 842
          :|||:|:||||::: |||| |::| ||||||| |:|| ::|:| ::||| |||||||:
Db    732 CCTTCTCATCCTGCATATGCCCCGCTGTTGGTCGTGTCGTCGGGTCCATGCGCCCGTACT 791

Qy    843 TGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGGTGGACATCTTGGCGGAGCAGG 902
          |||| |:| ||||| |:|| || :|:| |:|| ||| |:|:||| :||||| :||
Db    792 TGTTGGCCGTGCGTCGTCCATGGCCGGGCGCTGCTTGTGCGGGCCCATCATGTCCATGGCTGCT 851

Qy    903 AGTGGCAGGCGGCGCTCGTGCCCTTTAAGGTCATGAGCGGGGAAATGCCCTCCACCGAGGA 962
          ||| ||| || :| ::|:||| || :||:||:||:|: ::||| :|:|| |::||| ||||
Db    852 TGTGGCCGTCGTGCTTGTTGGCCGTGGTCCATGCGCCCGTACTTGTTGGCCATCATGTC 911

Qy    963 CCTGGTTAACTTACTCCCCTGCCATCCTCTCTCGGTGCCCCTGTGCGTCGTCGTGGGTCGGGGTCGTG 1020
          :|: :| :|| ::||:   ::|  ::||| |: ||||||: :|::| ::||:::|||:||:::||
Db    852 TGTGGCCGTCGTGCTTGTTGGCCGTGGTCCATGCGCCCGTACTTGTTGGCCATCATGTC 911
```

RESULT 6

```
Db  BX905657/c
912 CATGGCCGCCGTGCTTGTGTGGCCGTGCGTTCGTGCTTGTGTGGCCGTGCGTTCGTGCTTGTGTGGCCGTGCGTTCGTG 969
    ||||||  ||| ||  ||||  ||||  |||  ||||  ||||  |||  ||||  ||||  |||  ||||  ||||  ||||||
```

LOCUS   BX905657   635 bp   DNA   linear   GSS 07-JAN-2004
DEFINITION Leishmania braziliensis GSS, clone LBAF018C06, genomic survey
            sequence.
ACCESSION   BX905657
VERSION   BX905657.1 GI:40734125
KEYWORDS   GSS; genomic survey sequence.
SOURCE   Leishmania braziliensis
  ORGANISM Leishmania braziliensis
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania; Leishmania braziliensis species complex.
REFERENCE   1 (bases 1 to 635)
  AUTHORS   Laurentino,E.C., Ruiz,J.C. and Cruz,A.K.
  TITLE   GSS analysis of the Leishmania braziliensis genome
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 635)
  AUTHORS   Cruz,A.K.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-JAN-2004) Cruz A.K., University of Sao Paulo,
            Department of Molecular and Cell Biology, FMRP, Avenida
            Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
COMMENT   Clone requests: akcruz@fmrp.usp.br.
FEATURES   Location/Qualifiers
     source   1..635
                     /organism="Leishmania braziliensis"
                     /mol_type="genomic DNA"
                     /strain="MHOM/BR/75/M2904"
                     /db_xref="taxon:5660"
                     /clone="LBAF18C06"
ORIGIN

Query Match   1.3%; Score 47; DB 9; Length 635;
Best Local Similarity   48.7%; Pred. No. 1.2;
Matches 128; Conservative   0; Mismatches 135; Indels   0; Gaps   0;

```
Qy    780 CGCCCAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTTAGGCGCCGGCCATTGCTGGGTGC 839
          |||   ||   ||    |||     ||    |||||    ||   ||    |||   |||  |||
Db    382 CGCGCAGCTCCACGTTGCCGTTGCTGTTGACGCGCAGAGGTGCCACTGCGCACCGCTGCTGC 323

Qy    840 GGCTCGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGGGACATCTTGGCGGGGGTATGGGAGC 899
          ||    |||      |||    |||   |||   ||| ||||  || || ||||||  |||||
Db    322 TGCTCAGGAGGATCCATGACTTTCTCTCGGTAGAGATCGTAGTACTTCCACCTCGACGCCGA 263

Qy    900 AGGAGTGGGCGAGGGCCGCTCGTGGGCCTTTAAGGTCATGACGGCGAAAATGCCCTCCACCGA 959
          ||| || || || || |||||   || |||  ||| || || |||  ||||| ||   ||||||
Db    262 AGGAGTGGGCTGCGTGTCGTGTGCTCACGCTTCTGCGTGAGCTCCGCGAACAGATCCCCGCACGA 203

Qy    960 GGACCTGGTTAACTTACTCCCGTGCCCTGGTGCTCGTCGGGGGTCGT 1019
          |||   ||| || || ||  || |||  ||||| ||  ||||| ||||
Db    202 GGCCGGGTACAATGCCCTGAAGACCCTCTCGCTCCCATGTAGAGAGCCTTGCTGTGTCGT 143

Qy   1020 GTGCGCCAGCGATACTGCGTCGGC 1042
          |||   || ||    |||||| |||
Db    142 GTTCGGCCGCGCGTGTCGAGACTGC 120
```

RESULT 7

REFERENCE   1
  AUTHORS   Laurentino,E.C., Ruiz,J.C. and Cruz,A.K.
  TITLE   GSS analysis of the Leishmania braziliensis genome
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 635)
  AUTHORS   Cruz,A.K.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo,
            Department of Molecular and Cell Biology, FMRP, Avenida
            Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
COMMENT   Clone requests: akcruz@fmrp.usp.br.
FEATURES   Location/Qualifiers
     source   1..635
                     /organism="Leishmania braziliensis"
                     /mol_type="genomic DNA"
                     /strain="MHOM/BR/75/M2904"
                     /db_xref="taxon:5660"
                     /clone="LBAF018C06"
ORIGIN

Query Match   1.3%; Score 47; DB 9; Length 635;
Best Local Similarity   48.7%; Pred. No. 1.2;
Matches 128; Conservative   0; Mismatches 135; Indels   0; Gaps   0;

```
Qy    780 CGCCCAACTCGCTCCCCCCAGTGCTGCTTCAGCTTTCGTTAGGCGCCGGCCATTGCTGGGTGC 839
          |||   ||   ||    |||     ||    |||||    ||   ||    |||   |||  |||
Db    382 CGCGCAGCTCCACGTTGCCGTTGCTGTTGACGCGCAGAGGTGCCACTGCGCACCGCTGCTGC 323

Qy    840 GGCTCGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGGGACATCTTGGCGGGGGTATGGGAGC 899
          ||    |||      |||    |||   |||   ||| ||||  || || ||||||  |||||
Db    322 TGCTCAGGAGGATCCATGACTTTCTCTCGGTAGAGATCGTAGTACTTCCACCTCGACGCCGA 263

Qy    900 AGGAGTGGGCGAGGGCCGCTCGTGGGCCTTTAAGGTCATGACGGCGAAAATGCCCTCCACCGA 959
          ||| || || || || |||||   || |||  ||| || || |||  ||||| ||   ||||||
Db    262 AGGAGTGGGCTGCGTGTCGTGTGCTCACGCTTCTGCGTGAGCTCCGCGAACAGATCCCCGCACGA 203

Qy    960 GGACCTGGTTAACTTACTCCCGTGCCCTGGTGCTCGTCGGGGGTCGT 1019
          |||   ||| || || ||  || |||  ||||| ||  ||||| ||||
Db    202 GGCCGGGTACAATGCCCTGAAGACCCTCTCGCTCCCATGTAGAGAGCCTTGCTGTGTCGT 143

Qy   1020 GTGCGCCAGCGATACTGCGTCGGC 1042
          |||   || ||    |||||| |||
Db    142 GTTCGGCCGCGCGTGTCGAGACTGC 120
```

LOCUS   LBAF018C06   635 bp   DNA   linear   GSS 23-JUN-2003
DEFINITION Leishmania braziliensis GSS, clone LBAF018C06, genomic survey
            sequence.
ACCESSION   BX545484
VERSION   BX545484.1 GI:32169590
KEYWORDS   GSS; genomic survey sequence.
SOURCE   Leishmania braziliensis
  ORGANISM Leishmania braziliensis
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania; Leishmania braziliensis species complex.

RESULT 8

LOCUS   CO015154   855 bp   mRNA   linear   EST 09-JUN-2004
DEFINITION EST785536 Coccidioides posadasii saprobic phase cDNA library,
            greater than 4kb Coccidioides posadasii cDNA clone CIBBG75 5' end,
            mRNA sequence.
ACCESSION   CO015154
VERSION   CO015154.1 GI:48522043
KEYWORDS   EST.
SOURCE   Coccidioides posadasii
  ORGANISM Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1 (bases 1 to 855)
  AUTHORS   Gardner,M.J. and Cole,G.T.
  TITLE   Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
  JOURNAL   Unpublished (2003)
COMMENT   Other ESTs: EST785535
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..855

RESULT 9

```
LOCUS       CF823313               909 bp    mRNA    linear   EST 01-APR-2004
DEFINITION  Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDAK77 5' end, mRNA
            sequence.
ACCESSION   CF823313
VERSION     CF823313.1  GI:45923370
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
  ORGANISM  Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Gardner,M.J. and Cole,G.T.
  TITLE     Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
  JOURNAL   Unpublished (2003)
COMMENT     Other ESTs: EST700694
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..909
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDAK77"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN
```

Query Match   1.3%;   Score 47;   DB 7;   Length 909;
Best Local Similarity   52.9%;   Pred. No. 1.3;
Matches 101; Conservative   0; Mismatches   90; Indels   0; Gaps   0;

```
Qy        838 GCGGCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGGTGGACATCTTGGCGGGCTATGGA 897
              ||||  ||    |||||||||||||||||||||||||||| ||||||||||||||
Db        824 GTGCTTGTGGCCGTGCTTGTGTGGGCGTCGTCCATGGGCC  765
Qy        898 GCAGGAGTGGCAGGCGGCGCTCGTGGCCCTTTAAGGTCATGAGCGGCGAAATGCCCTCACC 957
              ||||      ||||||||||||||||||||||||||||||||||||||||||    ||||
Db        884 GTGCCTTGTGTCCATGGGCCTGTGTCGTGTCCATGGGCCGTGCTTGTGTGGGCC 825
Qy        958 GAGGACCTGGTTAACTTACTCCCTGCCATCCTCTCTCCTGGTGCCCTGGTCGTCGGGGTC 1017
              ||    |||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        764 ATGTCCATGGGCCGCGTGCTTGTGGGCCATCATGTCCATGGGCCGTGCTTGTGGGCCGTC 705
Qy       1018 GTGTGCGCAGC 1028
              ||||||
Db        704 GTGTCCATGGC 694
```

RESULT 10

```
LOCUS       CF826948               914 bp    mRNA    linear   EST 01-APR-2004
DEFINITION  Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDS573 5' end, mRNA
            sequence.
ACCESSION   CF826948
VERSION     CF826948.1  GI:45933005
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
  ORGANISM  Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Gardner,M.J. and Cole,G.T.
  TITLE     Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
  JOURNAL   Unpublished (2003)
COMMENT     Other ESTs: EST704329
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..914
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDS573"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN
```

Query Match   1.3%;   Score 47;   DB 7;   Length 914;
Best Local Similarity   52.9%;   Pred. No. 1.3;
Matches 101; Conservative   0; Mismatches   90; Indels   0; Gaps   0;

```
Qy        838 GCGGCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGGTGGACATCTTGGCGGGCTATGGA 897
              ||||  ||    |||||||||||||||||||||||||||| ||||||||||||||
```

RESULT 11

```
Qy    897  GTGCTTGTGTGGCGTCGTCCATGGCCGCGTCGTTGTGTGGCCATCATGTCCATGGCCGCC
Db    898  GCAGGAGTGGCAGCGCGGCGCGCTCGGTGGCCCTTTAAGTCATGAGGCGCGGAAATGCCCTCCACC  838
Db    837  GTGCTTGTGGCCGTCGTTGTGCCGTCGTTGTGGCCGTCGTTGTGCCGTCGTTGTGCCATC  778
Qy    958  GAGGACCTGGTTAACTTACTCCCTCCATCCATCTCTCTCCTGGTGCGCCGCTGGT        1017
Db    777  ATGTCCATGGCCGCGGCGCGCCGCTGTTGTGCGCCATCATGTCCATGGCCGCGGGGTC      718
Qy   1018  GTGTGCGCAGC                                                    1028
Db    717  GTGTCCATGGC                                                     707
```

RESULT 12

LOCUS   CF823378   923 bp   mRNA   linear   EST 01-APR-2004
DEFINITION EST700760 Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDAL17 5' end, mRNA
            sequence.
ACCESSION   CF823378
VERSION   CF823378.1 GI:4529435
KEYWORDS   EST.
SOURCE   Coccidioides posadasii
ORGANISM   Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1 (bases 1 to 923)
AUTHORS   Gardner,M.J. and Cole,G.T.
TITLE   Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
JOURNAL   Unpublished (2003)
COMMENT   Other ESTs: EST700759
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..923
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDAL17"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN
```

```
Query Match   52.9%;   Score 47;   DB 7;   Length 923;
Best Local Similarity    1.3%;   Pred. No. 1.3;
Matches  101;   Conservative   0;   Mismatches   90;   Indels   0;   Gaps   0;
```

```
Qy    838  GCGGCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGACATCTTGGCGGGCTATGGA  897
Db    907  GTGCTTGTGGCCGTCGTTGTGGTTGTCCATGGCCGCGTCGTTGTGCCGTCGTTGTGCCGC  848
Qy    898  GCAGGAGTGGCAGGCGCGCTCGGTGGCCCTTTAAGGTCATGAGGCGGCGAAATGCCCTCCACC  957
Db    847  GTGCTTGTGGCCGTCGTGTGCGCCGTCGTTGTGCCATGGCCATGGCCATGGCCATC  788
Qy    958  GAGGACCTGGTTAACTTACTCCCTCCATCCATCTCTCTCCTGGTGCGCCGCTGGTCGGGGTC  1017
Db    787  ATGTCCATGGCCGCGCCGCCGCCGTGCTTGTGGCCGTC                          728
```

RESULT 12

```
LOCUS       CF825990              925 bp    mRNA    linear    EST 01-APR-2004
DEFINITION  EST703372 Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDB026 5' end, mRNA
            sequence.
ACCESSION   CF825990
VERSION     CF825990.1  GI:45932047
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
ORGANISM    Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 925)
AUTHORS     Gardner,M.J. and Cole,G.T.
TITLE       Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
JOURNAL     Unpublished (2003)
COMMENT     Other ESTs: EST703371
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDB026"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN
```

```
Query Match   52.9%;   Score 47;   DB 7;   Length 925;
Best Local Similarity    1.3%;   Pred. No. 1.3;
Matches  101;   Conservative   0;   Mismatches   90;   Indels   0;   Gaps   0;
```

```
Qy    838  GCGGCTGTTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGACATCTTGGCGGGCTATGGA  897
Db    919  GTGCTTGTGGCCGTCGTTGTCCATGGCCGCGTCGTTGTGCCGTCGTTGTGCCGTCGTTGTGCCGC  860
Qy    898  GCAGGAGTGGCAGGCGCGCTCGGTGGCCCTTTAAGGTCATGAGGCGGCGAAATGCCCTCCACC  957
Db    859  GTGCTTGTGGCCGTCGTTGTGGCCGTCGTTGTCCATGGCCGCGTTGCCTTGGGGCCATC  800
Qy    958  GAGGACCTGGTTAACTTACTCCCTCCATCCATCTCTCTCCTGGTGCGCCGCTGGTCGGGGTC  1017
Db    799  ATGTCCATGGCGCGCCCGTGCTTGTGGCCGTCATCATGTCCATGGCCGCGTCGTTGTGGCCGTC  740
Qy   1018  GTGTGCGCAGC                                                    1028
Db    739  GTGTCCATGGC                                                     729
```

RESULT 13

```
LOCUS       CNS01213              645 bp    DNA    linear    GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit
```

fly), genomic survey sequence.

ACCESSION   AL101589
VERSION   AL101589.1 GI:5613200
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 645)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
             collaboration with the European Drosophila Genome Project (EDGP) -
             http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
             library (Dros BAC) was made by Alain Billaud at CEPH (Centre
             d'Etude du Polymorphisme Humain) with funding provided by a MRC
             project grant. The DNA was prepared from embryos by Alain Bucheton
             and Genevieve Payan. It has been constructed in the vector
             pBeloBAC11.
FEATURES   Location/Qualifiers
     source   1..645
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN08C07"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

Query Match   1.2%; Score 46.2; DB 9; Length 645;
Best Local Similarity   12.0%; Pred. No. 1.9;
Matches   55; Conservative 127; Mismatches 276; Indels   1; Gaps   1;

Qy   800 GTGCTGCTTCAGCTTTCGTAGGCGCCCGGCAATTGCTGGTGCGGCTCTTGGCAATAGGCC 859
            ::|||||:|: |::|:|||:::| |||:| |:||||||||||::||| |||:|:|
Db   129 KKGGNGGGNNNGTNGTGGTNGNNGGGGGGNNNNGGNNNNTTNNTNGGTTTNNNNNGGNNN 188

Qy   860 TTGGGAAGGTGCTTGTTGGACATCTTGGCGCGGGCTATGGACAGGAGTGGCCGCGCCTCG 919
            ::::::|| :|:|||:||:||:::||||||::|||:|||::|::|||:|||| ||:|||
Db   189 NNSCCCSNCTNCNNNNNTNTTNSSSSNSNNNTNNNNNNSSSSSSSSNSNNNNSSSSSNNN 248

Qy   920 TGGCCTTTAAGGTCATGACGGCGGAAATGCCCTCCACCGAGAGAGTGGTTAACTTACTCC 979
            :::|:::|:||:|| |||:||||| ::|::||::|:|||::|:|||:|:||::::::|||
Db   249 NSGGSSSNSNSNNGSGSSSSNSNTSSNNNNNNSSTNSNNNNSSTTSNSSSNNNNSSS 308

Qy   980 CTGCCATCCTCTCTCCTGGTGCCCTGGTCGTGTGTCGCCAGCGCAATACTGCGTC 1039
            ::|: ||:::|::|:|||:|||::|:|::||:||:: ::
Db   309 TSTTNTNSNTSSNNSATGSSNSSSGGTTBSGSSSNSSSNNNNNNASNNSBTSNS 368

Qy   1040 GGCACGTGGGTCCAGGGGAGGGGGGCTGTGCAGTGGATGGAACCGGCTGATAGCGTTCGCCT 1099
            ::||:|:::||:::|:::|||||||||::|||||||:::||:|::|::|::||:|||
Db   369 NNSNSATTSST-NCSSGSSSSSSSSSSSTSNSTSNSNSNSNNTSSSTTNSTST 427

Qy   1100 CGCGGGGGTTAACCATGTTTCCCCCACGCACTTATGTGCCAGAGAGGCACGACGCGCACACGTG 1159
            ::|||| |::|::||::|||:|||:|::|::|:|:||:|||:|:|:|:::|:::|::|:|:|
Db   428 SSSSSNSNSTGSSSSSTTTTNSSASSNNNNNGNSSSSSNAANSTSTTNSSNNTTTSSNNSNS 487

Qy   1160 TCACTCAGATCCTCCGACCTTACTATCACCCAACTGTTGAAGAGGCTCCACCAGTGGA 1219
            ::|::|::|:::|:|||::::|:|:|||::::|::|:||||:||:::|::|::|||
Db   488 SCNTTSANNTSSTTNTTTNTAGNNNNAATSSSNSNNSNNNSNSGGTSSGNTSSSSC 547

Qy   1220 TTAACGAGCAGACTGCTCACGCCCTGCTCCGGCTCGTGTGGC 1258
            ::|:||:||||: :|:::|:::|::|||:|:||:|:||:||::
Db   548 TGTBSSGNSNSNNCNGSNNSCTGNSNSNSNSSTTTGSS 586

LOCUS   CF821532   881 bp   mRNA   linear   EST 01-APR-2004
DEFINITION EST698914 Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii saprobic phase cDNA clone CIDAA65 5', end, mRNA
            sequence.
ACCESSION   CF821532
VERSION   CF821532.1 GI:45927589
KEYWORDS   EST.
SOURCE   Coccidioides posadasii
ORGANISM   Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1 (bases 1 to 881)
AUTHORS   Gardner,M.J. and Cole,G.T.
TITLE   Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
JOURNAL   Unpublished (2003)
COMMENT   Other_ESTs: EST698913
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..881
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDAA65"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN

Query Match   1.2%; Score 45.8; DB 7; Length 881;
Best Local Similarity   55.3%; Pred. No. 2.6;
Matches   89; Conservative   0; Mismatches 72; Indels   0; Gaps   0;

Qy   868 GTGCTTGTGGACATCTTGGCGCGGGCTATGGAGCAGGAGAGTGGCAGGGCGGCTCGTGGCCTTT 927
            |||||||||||| |||| ||||||| |||| |||| |||| ||||| |||||| ||||||||||||
Db   876 GTGCTTGTGGCCCATCATGTCCATGGCGCGCGCCGCGCGCTGCTTGGCCGTGCTTGGCGGTC 817

Qy   928 AAGGTCATGAGCGCGGGAAATGCCCTCCACCAGGACCTGGTTAACTTACTACTCCCCTCGCCATC 987
            |||||||||||| ||||||||||||| || || ||||| || ||| ||| ||||| |||||||
Db   816 GTGTCCATGGCCGCGCCGTCTTGTGCCGCCCATCATGTCCATGGCGCGCGCCGCGCGCTTGTGTGCCATC 757

Qy   988 CTCTCTCCTGGTGCCCTGGTCGTGTGTCGCCAGCGCAGC 1028
            ::|||| |:|||||| ||||||||| | ||||:|| ||||
Db   756 ATGTCCATGGCGCGCGCCGCGCGCTTGTGTGCCGTGCGTCGTCCATGGC 716

LOCUS   BP153627   506 bp   mRNA   linear   EST 30-DEC-2003
DEFINITION BP153627 full-length enriched swine cDNA library, adult ovary Sus
            scrofa cDNA clone OVRM10094G12 5', mRNA sequence.
ACCESSION   BP153627
VERSION   BP153627.1 GI:40403100
KEYWORDS   EST.
SOURCE   Sus scrofa (pig)
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 506)
AUTHORS   Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
            Okumura,N., Hamasima,N. and Awata,T.
TITLE   PEDE (Pig EST Data Explorer): construction of a database for ESTs

```
JOURNAL     Nucleic Acids Res. 32 (1), D484-D488 (2004)
COMMENT     Contact: Hirohide Uenishi
            Animal Genome Laboratory, Genome Research Department
            National Institute of Agrobiological Sciences
            2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
            Tel: +81-29-838-8627
            Fax: +81-29-838-8627
            Email: huenishi@affrc.go.jp
            This EST project with full-length enriched cDNA libraries carried out in
            Animal Genome Research Program (Japan) by National Institute of
            Agrobiological Sciences and STAFF-Institute
            Single pass sequencing of clones derived from oligo-capped cDNA
            library
            Vector sequences were eliminated by RepeatMasker version 2002/07/13
            and crossmatch version 0.990319
            Low quality bases were trimmed based on the quality values.
FEATURES             Location/Qualifiers
     source          1..506
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="OVRM10094G12"
                     /tissue_type="ovary"
                     /dev_stage="adult"
                     /clone_lib="full-length enriched swine cDNA library, adult
                     ovary"
ORIGIN
Query Match     1.2%;   Score 45.6;   DB 5;   Length 506;
Best Local Similarity  48.5%;   Pred. No. 2.6;
Matches 126;   Conservative 0;   Mismatches 134;   Indels 0;   Gaps 0;

Qy      2234 AGGCCCCGGACTACGTCCCTCCACAGTGGTACAATGGGTGCCCACTGCCACCTTACTAAGACCC 2293
             || ||| | |||| || ||||| ||||||||||| | |||||||||||||||||||||||||
Db       389 AGCCGCCTGGGCGTGGCTCCCTCCCAGCGCTCCTTGGCCACCTTAACTGCTCATCGATAAAACCG 330

Qy      2294 CTCCTATACCACCTTCCACCGGAGGAAGAGAGACAGTTGTTCTGACACAGAAATCCACCGTGTCTT 2353
             ||| | | || |||||||| | ||||| |||||| || |||||||||||||||||||||||
Db       329 TGCTGCTGCGGCCGCCGAACCGCGTCCGGGACGGAGGCCCAGGCCCCGGCGCCCGCCCTTTCCG 270

Qy      2354 CTGCCCTGGCGGGGAGCTTGCCACAAAGGCTTTCGGGTAGCTCCGAACCGTCGGCCGTCGACA 2413
             ||| ||| ||| ||||| ||||||
Db       269 CAGCGGCAGGCCGGAGCTCGCCGGGACCACCGGACCGGGCCGGGAGCAGGGATCTGACG 210

Qy      2414 GCGGCACGGCAACCGCCCCTCCTGACCAACCCTCCGACCGACGGCGGAGCAGGATCTGACG 2473
             ||| |||||| ||||| | |||| |||||||| | || |||||||||| ||||||||
Db       209 CGGGTCCGCCGACGAGGGCCTTGCAGGCCCCCCCACCCCGAAGGCGGGGGACCACCGGGTCCGCCG 150

Qy      2474 TTGAGTCGTATTCCTCCATG 2493
             ||| ||
Db       149 GGGGCGCGAGGGCCACCAGG 130
```

Search completed: February 21, 2005, 18:48:32
Job time : 11369 secs